特別研究論文

（査読済み）

研　究　題　目

調波打撃音モデルに基づく線形多チャネルブラインド音源分離

提出年月日	2022 年　1 月　31 日
氏　名	大藪　宗一郎
主　査	北村　大地　講師
副　査	村上　幸一　准教授
副　査	柿元　健　准教授

香川高等専門学校
専攻科
創造工学専攻

Linear Multichannel Blind Source Separation Based on Harmonic/Percussive Source Model

Soichiro Oyabu

Advanced Course in Industrial and Systems Engineering
National Institute of Technology, Kagawa College

Abstract

Blind source separation (BSS) is a technique for estimating specific audio sources from an observed mixture signal. Among them, BSS assuming a situation where the number of microphones is less than the number of sources is called underdetermined BSS. Also, BSS for a monaural observed signal is called single-channel BSS. For monaural signals, BSS is difficult because there is not enough information to find a solution, resulting in a generation of artificial distortions in the separated signals. On the other hand, BSS which assumes the same numbers of microphones and sources is called determined BSS. Determined BSS is a linear transformation, and this mechanism provides high-quality separation performance with minimum distortion. However, the performance of determined BSS strongly depends on fitness of a time-frequency assumption for each source (source model). Various source models including low-rankness and group sparsity have been assumed in determined BSS. In recent BSS studies, time-frequency-masking-based BSS (TFMBSS) was proposed, which enables us to flexibly design the source models via a time-frequency mask. In this thesis, I propose a new algorithm that achieves high-quality BSS of musical instruments and drums by utilizing HPSS to construct the source models (time-frequency masks) in TFMBSS. HPSS is popular single-channel BSS for separating harmonic and percussive sounds in music signals. The proposed algorithm achieves linear source separation by taking the HPSS-based source model into account. In addition, I propose a mask-smoothing method to stabilize the optimization algorithm of TFMBSS, whose efficacy is confirmed by the BSS experiments. The proposed method utilizes two typical single-channel HPSS algorithms. To compare the performances of conventional and proposed methods, I conducted the BSS experiments using professionally produced music signals. The experimental results show that the proposed method can provide more accurate separation of drums and the other harmonic sources compared with several conventional BSS algorithms.

Key Words: blind source separation (BSS), harmonic/percussive sound separation (HPSS), time-frequency masking, mask stabilization, plug-and-play scheme

（和訳）

ブラインド音源分離（BSS）とは，複数の音源が混合した観測信号から，混合前の音源を推定する技術である．その中でも，マイク数が音源数未満の状況を仮定する BSS を劣決定 BSS といい，特にマイク数が 1 つの場合を単一チャネル BSS という．単一チャネル BSS では，解を求めるための情報が少ないことから音源分離が困難である．そのため，分離音への人工歪みが発生してしまうことが問題となる．対して，音源数とマイク数が同じ状況を仮定する優決定 BSS は，線形変換であるため，分離音に歪みが生じにくく，高品質な音源分離が可能である．これら BSS の性能は，アルゴリズム内で仮定する各音源の時間周波数構造（音源モデル）の適合具合に強く依存する．特に優決定 BSS では，これまでに低ランク性や群スパース性等様々な音源モデルが仮定されてきたが，近年の研究で時間周波数マスクを音源モデルとして用いる優決定 BSS（TFMBSS）が提案され，柔軟な音源モデル構築が可能となった．本論文では，調波打撃音分離（HPSS）を TFMBSS の音源モデルの構築に活用することで，楽器音とドラム音の高品質な BSS を達成するアルゴリズムを新たに提案する．HPSS は音楽信号に含まれる調波音と打撃音の分離で有名な単一チャネル BSS である．このアルゴリズムにより，単一チャネル BSS の音源モデルに沿った分離を TFMBSS における線形分離で達成する．さらに，TFMBSS の最適化アルゴリズムを安定化させるためのスムージング法も提案する．実験では，提案したスムージング法の有効性を検証し，スムージング法の有用性を示した．さらに，これまでに提案されてきた代表的な 2 種類の単一チャネル HPSS を提案アルゴリズムに導入し，音源分離性能における比較実験を行った．実験結果として，いずれも提案アルゴリズムの有意性が十分に示される結果となった．

目次

第 1 章　緒言 1

1.1　音源分離の背景 1

1.2　本論文における主題 3

1.3　本論文における動機 4

1.4　本論文の構成 5

第 2 章　調波打撃音を対象とした従来の音源分離手法 6

2.1　まえがき 6

2.2　STFT 6

2.3　定式化 7

2.4　観測信号のチャネル数における音源分離手法の区分 9

2.5　HPSS 10

2.5.1　最適化に基づく HPSS 10

2.5.2　メディアンフィルタに基づく HPSS 11

2.5.3　多チャネル HPSS 12

2.6　TFMBSS 12

2.6.1　近接作用素 13

2.6.2　時間周波数マスク 13

2.6.3　主双対近接分離法 13

2.6.4　主双対近接分離法を応用した線形分離アルゴリズム 15

2.6.5　TFMBSS の概要 15

2.7　本章のまとめ 16

第 3 章　調波打撃音モデルに基づく線形多チャネル音源分離 17

3.1　まえがき 17

3.2　提案アルゴリズムの動機 17

3.3　提案アルゴリズムの概要 18

3.3.1　前後処理 18

3.3.2　OHPSS モデルにおける分離推定 21

3.3.3　MHPSS モデルにおける分離推定 22

3.3.4　時間周波数マスクの生成 22
3.4　マスクのスムージング 23
3.5　本章のまとめ 23

第 4 章　評価実験 25
4.1　まえがき 25
4.2　実験条件 25
4.3　OHPSS の反復回数における影響の検証 26
4.4　MHPSS のフィルタサイズにおける影響の検証 27
4.5　マスクのスムージングにおける影響の検証 28
4.6　OHPSS モデルと MHPSS モデルの性能比較実験 29
4.7　他の従来手法との性能比較実験 30
4.8　本章のまとめ 31

第 5 章　結言 34

謝辞 35

参考文献 35

付録 A　OHPSS モデルの各楽曲における SDR 改善量の収束挙動 42

付録 B　MHPSS モデルの各楽曲における SDR 改善量の収束挙動 53

付録 C　各楽曲における従来の BSS との性能比較 64

第 1 章

緒言

1.1 音源分離の背景

音源分離とは，観測したある混合音源から，混合前の信号を推定する技術である．具体的な利用例を Figs. 1.1 及び 1.2 に示す．まず，音源分離の例として音声信号に対する音源分離が挙げられる．音声信号に対する分離では，混合信号から雑音を除去して音声だけを抽出する雑音と音声の分離や，複数人が会話を行っている状況下で個人の音声毎に分離し抽出するような音声と音声の分離などがある．これらの研究の動機としては，スマートスピーカーやナビゲーションシステムなど音声認識技術を用いた応用システムが近年増えている中，雑音などが含まれる混合信号では入力信号の認識精度が著しく低下するという問題が存在し，雑音の混合がないクリアな音声音源が入力として求められていることが挙げられる．

もう一つの例として音楽信号に対する音源分離がある．音楽信号に対する分離では，ある観測した音楽信号から，ピアノ，ギター，ドラム，及びボーカルなどの音源毎に分離するという処理を行う．これらの研究の応用例として，近年，既存楽曲のオーディオの再編集を行うようなリミックス文化が形成されており，オーディオ編集を行うようなユーザは各楽器毎の高品質な分離音源を必要としていることが挙げられる．

上記のように，音源分離技術は近年ニーズが高まっており，これらのタスクを満足するには高精度な音源分離手法が求められる．この経緯から，1990 年代から今日まであらゆる音源分離手法が提案されてきた．その音源分離手法の中でも，マイクロホンや音源の位置等の事前情報を用いずに，複数の音源が混合した観測信号から，混合前の音源信号を推定する技術をブラインド音源分離（blind source separation: BSS）という [1, 2]．BSS の概要を Fig. 1.3 に示す．未知の混合系 $\boldsymbol{A}$（マイクロホンや音源位置や部屋の形状などに依存して変化）から混合信号が生成される．これに対して混合系 $\boldsymbol{A}$ の逆系である分離系 $\boldsymbol{W}$ を推定し混合信号に適用することで混合前の音源を推定するという仕組みである．

観測マイクロホン数が元の音源数以上となる優決定条件下での BSS には，独立成分分析（independent component analysis: ICA）[3] に基づく手法が広く用いられている．例えば，ICA を周波数毎に適用した周波数領域 ICA（frequency-domain ICA: FDICA）[4]–[6] や，FDICA におけるパーミュテーション問題 [7] の解決と分離行列の推定を同時に行う独

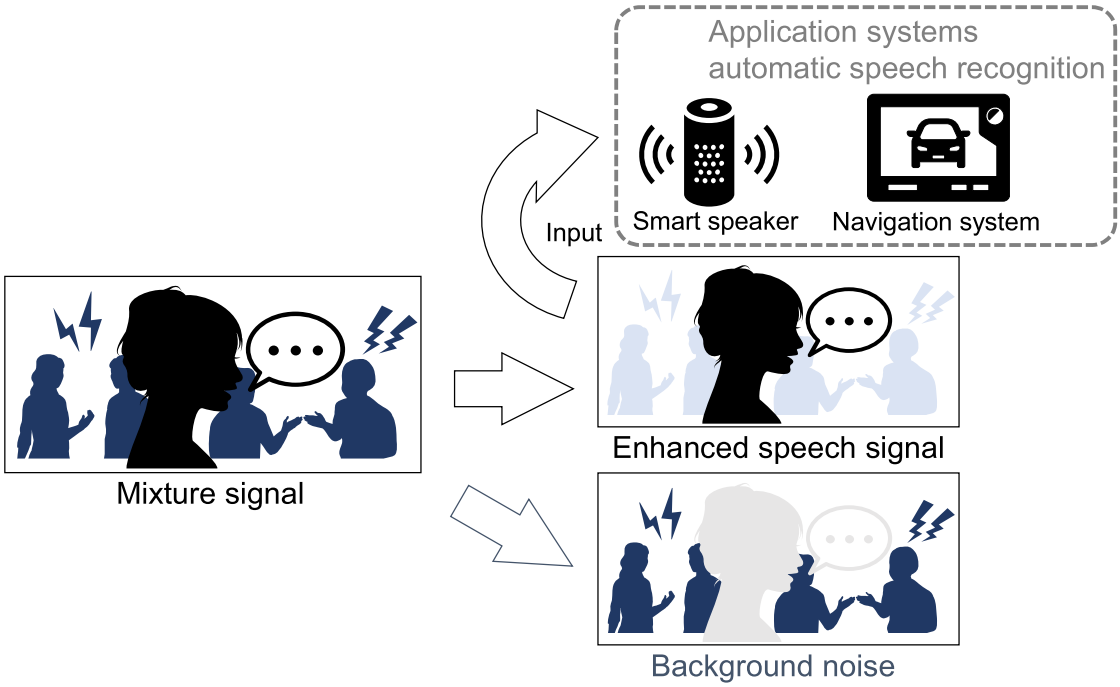

Fig. 1.1. Application example of speech source separation.

立ベクトル分析（independent vector analysis: IVA）[8]–[10] 及び独立低ランク行列分析（independent low-rank matrix analysis: ILRMA）[11, 12] 等が提案されている．

特に高精度な音源分離が可能な優決定 BSS の IVA 及び ILRMA は，音源信号に関する事前知識（音源モデル）に基づいてパーミュテーション問題を解決しており，仮定する音源モデルと実際の混合信号との適合度合いによって，性能が大きく依存する．例えば，IVA は同一音源の全周波数成分が同時に強いパワーを持つことを仮定しており，これは音声信号の時間周波数構造に良く適合する音源モデルである．ILRMA は非負値行列因子分解（nonnegative matrix factorization: NMF）[13]–[15] を用いることで，同一音源の時間周波数構造が低ランク行列で高精度に近似できることを仮定しており，ドラムや楽器音などの音楽信号の時間周波数構造に良く適合する音源モデルである．これらの手法より分かることは，より良い音源モデルを BSS に導入できれば，より高品質な分離信号が得られる可能性があるということである．そのため，種々の音源モデルを用いて BSS の性能を比較することが重要と言える．

この目的に対して，主双対近接分離法 [16]–[20] を用いて幅広い音源モデルを統一的に扱える優決定 BSS アルゴリズムが提案された [21]．この手法では，近接作用素 [20] が計算できる関数で表される音源モデルであれば，容易に優決定 BSS に導入することができる．そして，この近接作用素は多くの有用な音源モデルにおいて閾値処理として与えることができ，時間周波数マスキングとして再解釈可能である．この解釈に基づく BSS として，時間周波数マスキングに基づく優決定 BSS（time-frequency-masking-based determined BSS: TFMBSS）[22, 23] が提案された．TFMBSS は，時間周波数マスクで表される音源モデルを用いて，線形の（歪みの少ない）多チャネル BSS が可能である．TFMBSS の応用例として IVA の音源モ

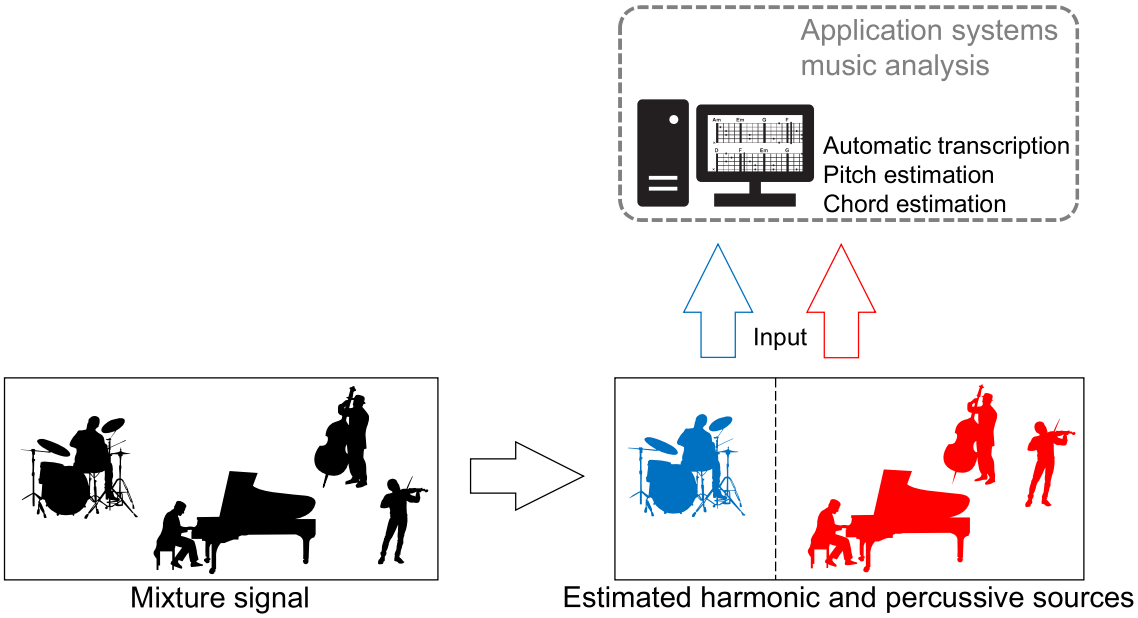

Fig. 1.2. Application example of music source separation.

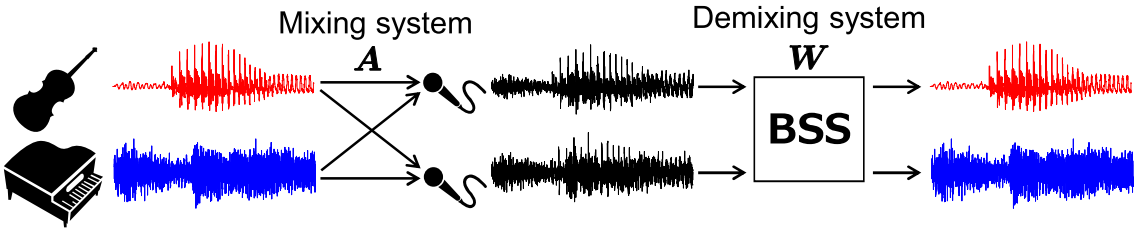

Fig. 1.3. Overview of BSS.

デルにスパース性を追加したスパース IVA [10]，ILRMA の音源モデルにスパース性を追加したスパース ILRMA [22]，及び調波構造の音源モデルを仮定した調波ベクトル分析（harmonic vector anarisys: HVA）[23] などの効果が検証されている．なお，TFMBSS と類似する手法として，補助関数に基づく IVA の分散に時間周波数パワーの推定値を用いるモデルベース IVA [24] が提案されているが，TFMBSS は (a) 最適化に近接分離法を用いる点及び (b) 独立性最大化という統計的枠組みを超える点の 2 点で大きく異なる．

1.2 本論文における主題

本論文では，TFMBSS の「柔軟な音源モデルを構築・活用できる」という利点を活かし，音楽信号の分離で有名な調波打撃音分離（harmonic/percussive source separation: HPSS [25]–[30] を TFMBSS の音源モデルに導入することを提案する．提案アルゴリズムは，調波音（ピアノやギターなどの音階を持った楽器音）と打撃音（ドラム音）を対象とした高品質な BSS

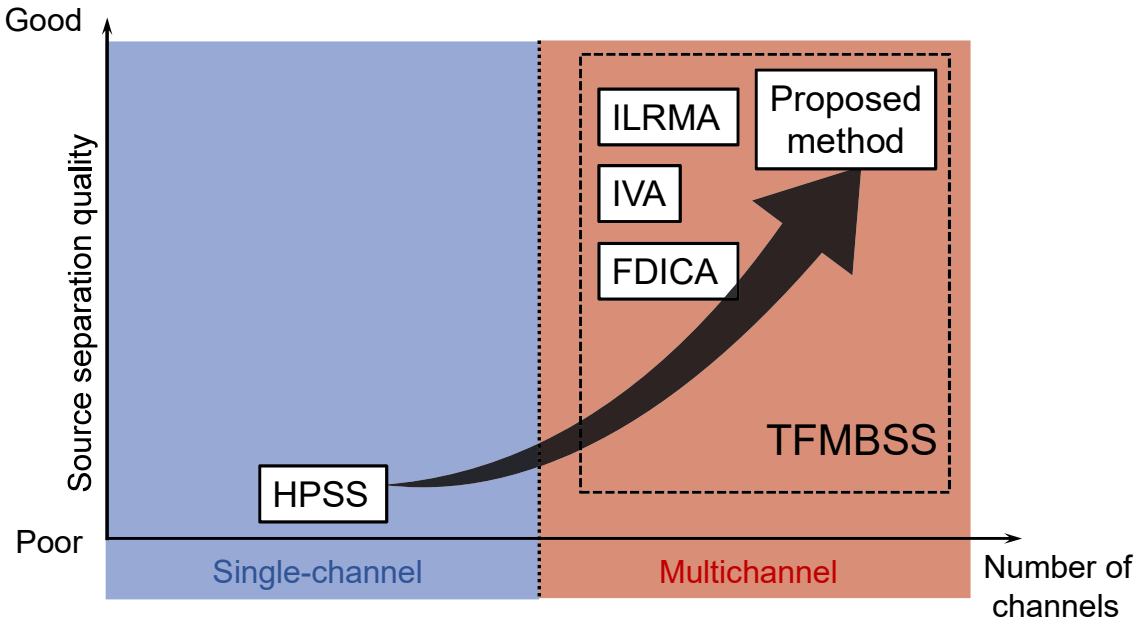

Fig. 1.4. Scope of this thesis.

として活用できる．提案アルゴリズムの有用性については次節で述べる．また，提案アルゴリズムの各パラメータについて，多数の楽曲を用いた実験的な調査も行う．

この提案アルゴリズムの既存手法に対する立ち位置を Fig. 1.4 に示す．HPSS は，複数の音源信号が混合したモノラル信号を対象とする単一チャネル BSS である．この単一チャネル BSS では，音源を分離するために非線形な処理が必要となり，その結果分離音に人工歪みが発生してしまう．音楽信号においては，このような人工歪みが発生した場合，芸術性が著しく損なわれてしまうことになる．提案アルゴリズムでは，この問題を解決するために，優決定 BSS の線形分離を活用しながら，HPSS の音源モデルに沿った分離を行う．さらに，これまでに提案されてきた代表的な 2 種類の単一チャネル HPSS のいずれかを選択して提案アルゴリズムに導入し，音源モデルの精度に対する性能比較を行う．加えて，既存の主要な優決定 BSS とも比較することで，提案アルゴリズムの有用性を確認する．

1.3　本論文における動機

音楽信号を調波音と打撃音に分離することは，多くの音楽信号処理手法の前処理として有用性が認められている [31]．例えば，リズムマップの自動生成手法への応用 [32]，楽曲における音階推定の前処理としての導入 [33, 34]，音楽信号の和音抽出技術の性能向上 [35] などが挙げられる．さらに，これらの音楽信号の解析（コード・テンポ・音階等の推定）を用いた，音楽情報検索（music information retrieval: MIR）が学術的な分野として近年発展している [36]．MIR は音楽を入力媒体とし，入力された音楽の特徴や性質を解析することで，関連する音楽を検索するようなシステムである．MIR の解析工程において，HPSS のような分離は重

要な役割を持っている．MIR の研究分野におけるコミュニティとして，音楽情報検索コンテスト（music information retrieval evaluation exchange: MIREX）[37] も年々活発化している．これらのことから，調波音と打撃音に関する分離手法の性能向上は，上記の技術の精度向上を大きく助長すると言える．

1.4 本論文の構成

本節にて，本論文の構成を示す．まず，2 章では，音響信号処理でよく用いられる基本的な信号の変換について説明し，さらに本論文における関連の既存手法として 3 種類の HPSS 及び TFMBSS の概要について述べる．3 章では，本論文の提案アルゴリズムについて，ステップ毎に詳細に述べる．4 章では，提案アルゴリズムのパラメータ探索実験，2 種類の音源モデルに対する比較実験，及び既存の BSS との比較実験を行い分離性能を評価する．最後に 5 章では, 全ての章を総括した結言を述べる．

第 2 章

調波打撃音を対象とした従来の音源分離手法

2.1　まえがき

本章では，本論文における重要な基礎知識を記載する．まず，2.2 節では，音響信号処理において重要な前処理となる短時間フーリエ変換（short-time Fourier transform: STFT）について解説する．2.3 節では，時間周波数領域における音源信号及び BSS の定式化を行う．2.4 節では，観測のマイクロホンによって分類される BSS について解説する．2.5 節では，調波音と打撃音の分離に用いられる有名な 3 種類の HPSS を紹介し，2.6 節では，TFMBSS の歴史と概要について詳細に解説する．2.7 節では，本章におけるまとめを述べる．

2.2　STFT

STFT とは，1 次元の時間信号を 2 次元の時間周波数信号に変換する処理である．STFT の処理の概要を Fig. 2.1 に示す．STFT の分析窓関数の長さ及びシフト長をそれぞれ Q 及び τ としたとき，時間領域の信号 $z[l]$ の j 番目の短時間区間信号（時間フレーム）は次式で表せられる．

$$\begin{aligned} \boldsymbol{z}^{(j)} &= [z[(j-1)\tau+1], z[(j-1)\tau+2], \cdots, z[(j-1)\tau+Q]]^T \\ &= \left[z^{(j)}[1], z^{(j)}[2], \cdots, z^{(j)}[q], \cdots, z^{(j)}[Q]\right]^T \in \mathbb{R}^Q \end{aligned} \tag{2.1}$$

ここで，$l=1,2,\cdots,L$, $j=1,2,\cdots,J$，及び $q=1,2,\cdots,Q$ はそれぞれ離散時間のインデクス，時間フレーム，及び時間フレーム内のサンプルを表し，$\cdot^{\mathrm{T}}$ は転置，$\mathbb{R}$ は実数全体の集合である．また，時間フレーム数 J は次式により与えられる．

$$J = \frac{L}{\tau} \tag{2.2}$$

但し，信号長 L はセグメント数 J が整数となるように各時間フレームの信号の両端にゼロを挿入する処理（ゼロパディング）が施される．また，各時間フレームにおける STFT は次式

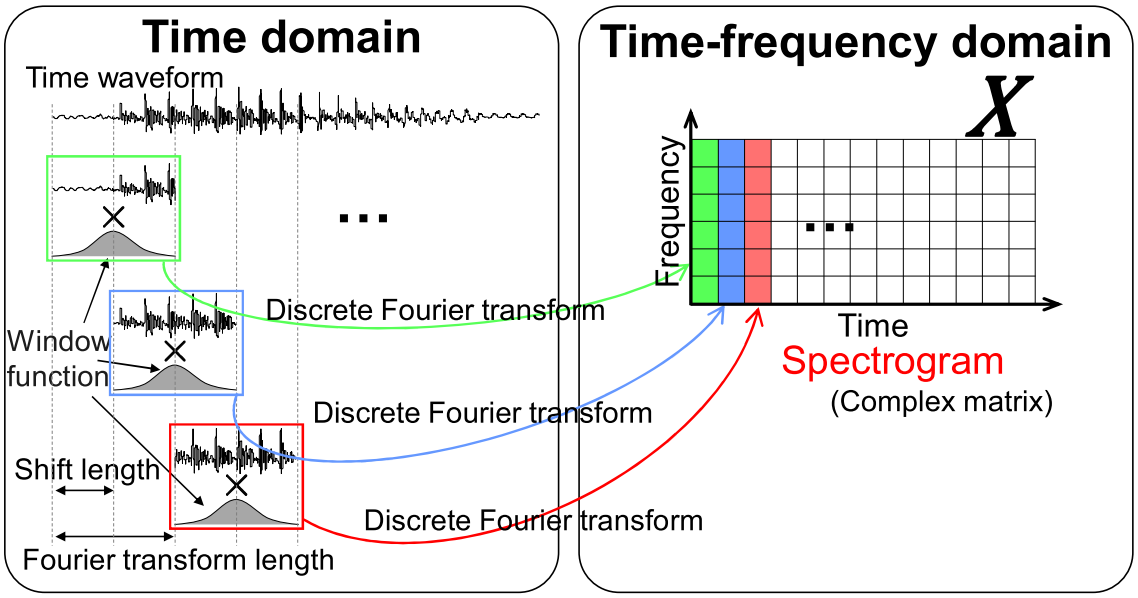

Fig. 2.1. Overview of STFT.

のように求められる．

$$\boldsymbol{Z} = \mathrm{STFT}_{\boldsymbol{\omega}}(\boldsymbol{z}^{(1)}, \boldsymbol{z}^{(2)}, \cdots, \boldsymbol{z}^{(J)}) \in \mathbb{C}^{I \times J} \tag{2.3}$$

ここで，$\mathbb{C}$ は複素数全体の集合である．スペクトログラム $\boldsymbol{Z}$ の (i,j) 番目の要素は次式で表される．

$$z_{ij} = \sum_{q=1}^{Q} \omega[q] z^{(j)}[q] \exp\left\{\frac{-\iota 2\pi (q-1)(i-1)}{\mathrm{T}}\right\} \tag{2.4}$$

ここで，F は $\left\lfloor \frac{F}{2} \right\rfloor + 1 = I$ を満たす整数（$\lfloor \cdot \rfloor$ は床関数），$i = 1, 2, \cdots, I$ は周波数ビンのインデクス，ι は虚数単位，$\boldsymbol{\omega} = [\omega[1], \omega[2], \cdots, \omega[L]]^{\mathrm{T}} \in \mathbb{R}^{Q}$ は分析窓関数をそれぞれ表す．この窓関数をを信号 $\boldsymbol{z}^{(j)}$ に乗じることで，両端の不連続性をスムーズにしている．以上の処理から，時間と周波数両方の情報を持った複素行列 $\boldsymbol{Z}$ が得られる．これをスペクトログラムと呼ぶ．音響信号処理においては，このスペクトログラム $\boldsymbol{Z}$ を信号処理の対象とするのが一般的である．実際の音声信号及び音楽信号のスペクトログラムを Fig. 2.2 に示す．

2.3 定式化

音源数と観測チャネル数（マイクロホン数）をそれぞれ N 及び M とし，多チャネルの時間信号を STFT して得られる時間周波数毎の音源信号，観測信号，及び分離信号をそれぞれ

$$\boldsymbol{s}_{ij} = [s_{ij1},\ s_{ij2},\ \cdots, s_{ijn},\ \cdots,\ s_{ijN}]^{\mathrm{T}} \in \mathbb{C}^{N} \tag{2.5}$$

$$\boldsymbol{x}_{ij} = [x_{ij1},\ x_{ij2},\ \cdots, x_{ijm},\ \cdots,\ x_{ijM}]^{\mathrm{T}} \in \mathbb{C}^{M} \tag{2.6}$$

$$\boldsymbol{y}_{ij} = [y_{ij1},\ y_{ij2},\ \cdots, y_{ijn},\ \cdots,\ y_{ijN}]^{\mathrm{T}} \in \mathbb{C}^{N} \tag{2.7}$$

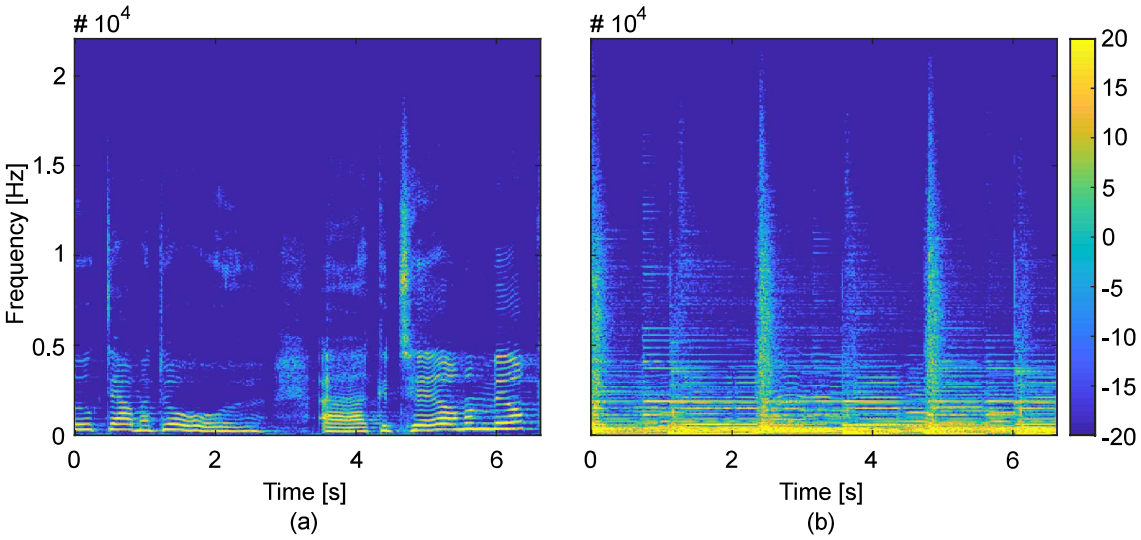

Fig. 2.2. Spectrograms represented by color map: (a) speech and (b) music signals.

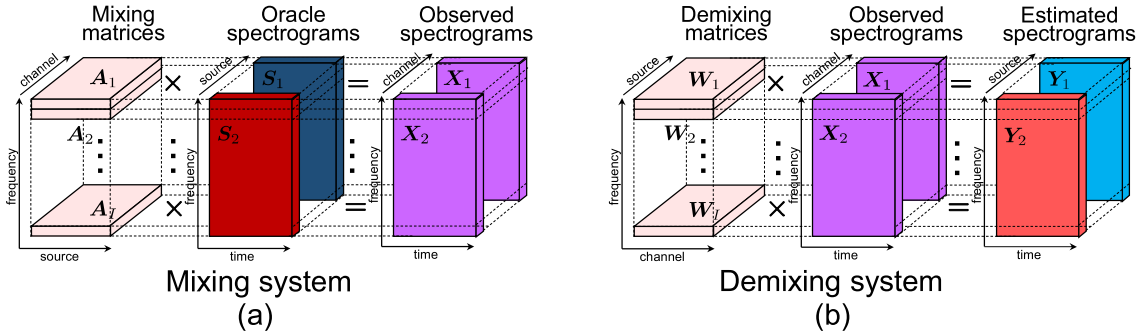

Fig. 2.3. Schematic explanation of (a) mixing and (b) demixing systems.

と表す．ここで，$n=1,2,\cdots,N$ は音源インデクス及び $m=1,2,\cdots,M$ はチャネルインデクスをそれぞれ示す．また，各信号の複素スペクトログラムを $\boldsymbol{S}_n \in \mathbb{C}^{I\times J}$，$\boldsymbol{X}_m \in \mathbb{C}^{I\times J}$，及び $\boldsymbol{Y}_n \in \mathbb{C}^{I\times J}$ で表す．これらの行列の要素はそれぞれ s_{ijn}，x_{ijm}，及び y_{ijn} である．

音源信号 $\boldsymbol{s}_{ij}$，観測信号 $\boldsymbol{x}_{ij}$，及び分離信号 $\boldsymbol{y}_{ij}$ の模式図を Fig. 2.3 に示す．混合系が線形時不変であり，時間周波数領域での複素瞬時混合で表現できると仮定すると，周波数毎の時不変な複素混合行列 $\boldsymbol{A}_i=[\boldsymbol{a}_{i1}\ \boldsymbol{a}_{i2}\ \cdots\ \boldsymbol{a}_{iN}]\in\mathbb{C}^{M\times N}$（ここで $\boldsymbol{a}_{in}=[a_{in1},\ a_{in2},\ \cdots,\ a_{inM}]^{\mathrm{T}}$ は各音源のステアリングベクトル）が定義でき，観測信号と音源信号の関係を次式で表現できる．

$$\begin{aligned}\boldsymbol{x}_{ij} &= \boldsymbol{A}_i\boldsymbol{s}_{ij}\\ &= \begin{bmatrix} a_{i11} & \cdots & a_{iN1} \\ \vdots & & \vdots \\ a_{i1M} & \cdots & a_{iNM} \end{bmatrix} \begin{bmatrix} s_{ij1} \\ \vdots \\ s_{ijM} \end{bmatrix}\end{aligned} \tag{2.8}$$

ここで，$\boldsymbol{A}_i$ は部屋の形状，マイクロホンの位置，及び部屋の残響度合いなどの情報から成る線形システムである．この混合モデルは，時不変混合系の残響時間が STFT の窓長よりも十

分短い場合に近似的に成立する．このとき，$M=N$ かつ $\boldsymbol{A}_i$ が正則であれば，分離ベクトル $\boldsymbol{w}_{in}=[w_{in1},\ w_{in2},\ \cdots,\ w_{inM}]^{\mathrm{T}}$ で構成される分離行列 $\boldsymbol{A}_i^{-1}=\boldsymbol{W}_i=[\boldsymbol{w}_{i1}\ \boldsymbol{w}_{i2}\ \cdots\ \boldsymbol{w}_{iN}]^{\mathrm{H}}\in\mathbb{C}^{N\times N}$ が存在し，分離信号は次式で与えられる．

$$\begin{aligned}\boldsymbol{y}_{ij}&=\boldsymbol{W}_i\boldsymbol{x}_{ij}\\&=\begin{bmatrix}\overline{w}_{i11}&\cdots&\overline{w}_{i1M}\\\vdots&&\vdots\\\overline{w}_{iN1}&\cdots&\overline{w}_{iNM}\end{bmatrix}\begin{bmatrix}x_{ij1}\\\vdots\\x_{ijM}\end{bmatrix}\end{aligned}\tag{2.9}$$

ここで，$.^{\mathrm{H}}$ はエルミート転置，$\overline{\cdot}$ は複素共役を示す．優決定 BSS では，式 (2.9) 中の分離行列 $\boldsymbol{W}_i$ を全周波数（$i=1,2,\cdots,I$）において推定することが最終的な目標となる．分離信号 $\boldsymbol{y}_{ij}$ は，観測信号 $\boldsymbol{x}_{ij}$ に対して周波数毎の分離行列を乗じることで推定されるため，式 (2.9) による音源分離は線形フィルタによる処理と等価であり，人工的な歪みが少ない（自然性の高い）分離が可能である．但し，$\boldsymbol{A}_i^{-1}=\boldsymbol{W}_i$ となる分離行列は $N=M$ の条件を満たさなければ存在しないため，式 (2.9) の音源分離は優決定条件（$N\leq M$）でのみ適用可能である．（$N<M$ では主成分分析等の次元圧縮を適用して $N=M$ とすることが一般的である）．

2.4　観測信号のチャネル数における音源分離手法の区分

音源分離に用いられる技術は，分離対象の観測チャネル数の違いよって明確な区分が存在する．その概要を Fig. 2.4 に示す．観測チャネル数が一つの場合，単一チャネル（モノラル）音源分離という．例として，HPSS [25]–[30] や半教師あり NMF [38]–[40] が挙げられる．また，近年では単一チャネル音源分離として，深層ニューラルネットワーク（deep neural network: DNN）を活用した音源分離手法が非常に多く提案されている [41]–[45].

一方，複数の観測チャネルで録音された音響信号を対象とする場合は，観測チャネル数が音源数より少ない場合（$N>M$）と，観測チャネル数が音源数以上の場合（$N\leq M$）が考えられる．前者は劣決定条件，後者は優決定条件と呼ばれる．特に優決定条件の音源分離は，式 (2.9) のように線形分離フィルタ $\boldsymbol{W}_i$ が構成できるため，歪みの少ない高品質な音源分離が達成できる可能性がある．優決定条件の音源分離の例としては，ICA [4]–[6]，IVA [8],–[10]，ILRMA [11, 12] などが挙げられる．

観測チャネル数が少ないということは，解を求めるための情報量が少ないということである．そのため，観測チャネル数が少ないほど音源分離は困難であり，分離音の音質が劣化する傾向にある．従って，芸術性が重要となる音楽信号などでは，低品質になりがちな単一チャネル音源分離や劣決定音源分離を適用しても，その後のアプリケーションに活用できない可能性がある．加えて DNN を活用した音源分離手法は，既存の非線形音源分離手法に比べて大幅な性能向上を見せているが，DNN は非線形写像であるため，非線形処理に起因する人工歪みの低減は保障されていないという欠点もある．以上の議論より本論文では，可能な限り高品質な（芸術性を失わない程の）音源分離を達成することを目的としているため，優決定 BSS を対象とする．

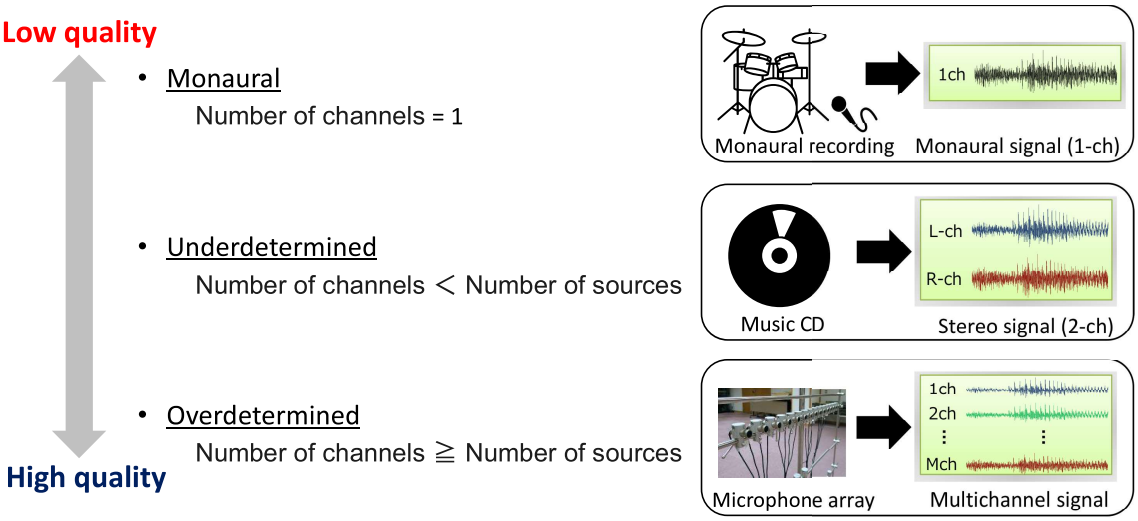

Fig. 2.4. Classification of audio source separation techniques based on number of channels of observed signals.

2.5　HPSS

HPSS は調波楽器及び打楽器の音源が持つ，対極的な振幅スペクトログラムの構造を仮定とした音源モデルに基づいて，次式のように，混合信号を調波音及び打撃音に分離する手法である．

$$\boldsymbol{B} = \boldsymbol{H} + \boldsymbol{P} \tag{2.10}$$

ここで，$\boldsymbol{B} \in \mathbb{C}^{I \times J}$，$\boldsymbol{H} \in \mathbb{C}^{I \times J}$，及び $\boldsymbol{P} \in \mathbb{C}^{I \times J}$ は，それぞれモノラルの混合信号の複素スペクトログラム，分離された調波信号の複素スペクトログラム，及び分離された打撃信号の複素スペクトログラムである．HPSS の概要を Fig. 2.5 に示す．HPSS は，調波音は時間方向に連続（Fig. 2.5 中の赤矢印）であり，打撃音は非定常的でかつ周波数方向に連続（Fig. 2.5 中の青矢印）である，という振幅スペクトログラム構造の特徴に着目している．本節では，3 種類の代表的な HPSS を従来手法として説明する．

2.5.1　最適化に基づく HPSS

有名な HPSS の一つに最適化に基づく HPSS（optimization-based HPSS: OHPSS）[25]–[29] が提案されている．OHPSS では，前節で説明した $\boldsymbol{H}$ と $\boldsymbol{P}$ の構造の違いを目的関数の最適化で表現し，この最適化問題を解くことで打撃音及び調波音を推定する．

文献 [25] の HPSS では，混合信号 $\boldsymbol{B}$ から $\boldsymbol{H}$ と $\boldsymbol{P}$ を推定するために，次式の目的関数で最

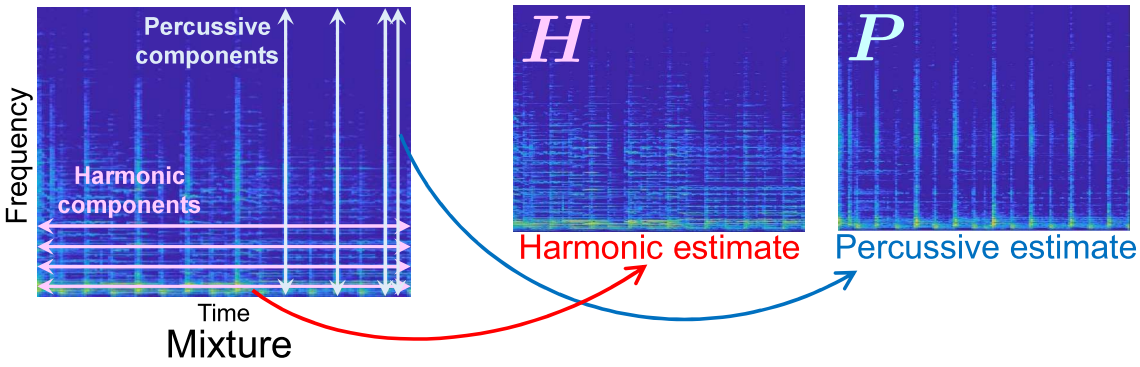

Fig. 2.5. Separation principle of HPSS.

小化する $\boldsymbol{H}$ 及び $\boldsymbol{P}$ を推定する．

$$J(\boldsymbol{H},\boldsymbol{P}) = \sum_{i,j} \left\{ \gamma_H \left(|h_{i(j+1)}|^{0.5} - |h_{ij}|^{0.5} \right)^2 + \gamma_P \left(|p_{(i+1)j}|^{0.5} - |p_{ij}|^{0.5} \right)^2 \right\} \tag{2.11}$$

ここで，h_{ij} 及び p_{ij} はそれぞれ $\boldsymbol{H}$ 及び $\boldsymbol{P}$ の要素であり，$\gamma_H > 0$ 及び $\gamma_P > 0$ は各項への重み係数である．なお，式 (2.11) の最小化問題においては，次に示される拘束条件が課せられている．

$$|b_{ij}| = |h_{ij}| + |p_{ij}| \tag{2.12}$$

$$\arg b_{ij} = \arg h_{ij} = \arg p_{ij} \tag{2.13}$$

式 (2.11) の最小化する h_{ij} 及び p_{ij} は，次式の反復更新式を全ての i 及び j について繰り返し計算することで推定できる [25]．

$$|h_{ij}|^{0.5} = \frac{\gamma_H \left(|h_{(i+1)j}|^{0.5} + |h_{(i-1)j}|^{0.5} \right) |b_{ij}|^{0.5}}{\sqrt{{\gamma_H}^2 \left(|h_{(i+1)j}|^{0.5} + |h_{(i-1)j}|^{0.5} \right)^2 + {\gamma_P}^2 \left(|p_{i(j+1)}|^{0.5} + |p_{i(j-1)}|^{0.5} \right)^2}} \tag{2.14}$$

$$|p_{ij}|^{0.5} = \frac{\gamma_P \left(|p_{i(j+1)}|^{0.5} + |p_{i(j-1)}|^{0.5} \right) |b_{ij}|^{0.5}}{\sqrt{{\gamma_H}^2 \left(|h_{(i+1)j}|^{0.5} + |h_{(i-1)j}|^{0.5} \right)^2 + {\gamma_P}^2 \left(|p_{i(j+1)}|^{0.5} + |p_{i(j-1)}|^{0.5} \right)^2}} \tag{2.15}$$

2.5.2 メディアンフィルタに基づく HPSS

OHPSS とは異なるアルゴリズムで，調波音及び打撃音のスペクトログラムを推定する技術として，メディアンフィルタに基づく HPSS（median-filter-based HPSS: MHPSS）[30] が提案されている．MHPSS は，振幅スペクトログラムの時間方向及び周波数方向にそれぞれメディアンフィルタを適用する．メディアンフィルタは，フィルタを適用する方向のスパイク状の成分を除去できるため，非線形かつ強力な平滑化が各方向に施される．従って，時間方向及び周波数方向の滑らかさを強調した信号を推定することができ，調波音及び打撃音が得られる．

MHPSS では，フィルタサイズ $2D+1$ の移動メディアンフィルタをシフト長 1 点でずらしながら適用する．メディアンフィルタを適用するベクトルは，次式のように混合信号の振幅スペクトログラム $|\boldsymbol{B}|$ の行及び列ベクトルとなる．

$$\boldsymbol{b}_{ij}^{(\mathrm{r})} = \left[|b_{i(j-D)}|, |b_{i(j-D+1)}|, \cdots, |b_{ij}|, |b_{i(j+1)}|, \cdots, |b_{i(j+D)}|\right] \in \mathbb{R}_{\geq 0}^{2D+1} \tag{2.16}$$

$$\boldsymbol{b}_{ij}^{(\mathrm{c})} = \left[|b_{(i-D)j}|, |b_{(i-D+1)j}|, \cdots, |b_{ij}|, |b_{(i+1)j}|, \cdots, |b_{(i+D)j}|\right] \in \mathbb{R}_{\geq 0}^{2D+1} \tag{2.17}$$

これらのベクトルにメディアンフィルタを適用することで，h_{ij} 及び p_{ij} が推定できる．

$$|h_{ij}| = \mathrm{median}(\boldsymbol{b}_{ij}^{(\mathrm{r})}) \tag{2.18}$$

$$|p_{ij}| = \mathrm{median}(\boldsymbol{b}_{ij}^{(\mathrm{c})}) \tag{2.19}$$

ここで，median$(\cdot)$ は入力されたベクトルの中央値のみをスカラーとして返す関数である．

2.5.3　多チャネル HPSS

本論文と同様の動機で，HPSS の多チャネル手法化を目指した多チャネル HPSS [46] が提案されている．多チャネル HPSS は，単一チャネル HPSS で仮定される音源モデルに加えて空間共分散行列モデル（spatial covariance model: SCM）[47, 48] を用いることで，空間的な伝達系に関するモデルを仮定している．多チャネル HPSS では，式 (2.10) の混合音及び分離音を多次元複素 Gauss 分布の確率変数としてモデル化し，その際の空間相関行列を次式で表現する．

$$\boldsymbol{\Sigma}_{ij}^{(\boldsymbol{B})} = v_{ij}^{(\boldsymbol{H})} \boldsymbol{R}_j^{(\boldsymbol{H})} + v_{ij}^{(\boldsymbol{P})} \boldsymbol{R}_j^{(\boldsymbol{P})} \tag{2.20}$$

ここで，$v_{ij}^{(\boldsymbol{H})}, v_{ij}^{(\boldsymbol{P})} \in \mathbb{C}$ は調波信号及び打撃信号のパワースペクトログラムであり，$\boldsymbol{R}_j^{(\boldsymbol{H})}, \boldsymbol{R}_j^{(\boldsymbol{P})} \in \mathbb{C}^{N\times N}$ は，各音源信号の空間的な広がり（伝達系）を示す時不変な SCM である．また，$\boldsymbol{\Sigma}_{ij}^{(\boldsymbol{B})} \in \mathbb{C}^{N\times N}$ は混合音の SCM を表す．そして，上式に対する対数尤度が次式で与えられている．

$$\log \mathcal{L} = -\sum_{i,j} \mathrm{tr}\left(\boldsymbol{\Sigma}_{ij}^{(\boldsymbol{B})-1} \widehat{\boldsymbol{\Sigma}}_{ij}^{(\boldsymbol{B})}\right) + \log\det\left(\pi \boldsymbol{\Sigma}_{ij}^{(\boldsymbol{B})}\right) \tag{2.21}$$

ここで，det$(\cdot)$ は行列式を表し，tr$(\cdot)$ は正方行列に対するトレースを表す．$\widehat{\boldsymbol{\Sigma}}_{ij}^{(\boldsymbol{B})}$ は混合音の SCM の推定成分である（定義の詳細は文献 [48] を参照）．

この時，$v_{ij}^{(\boldsymbol{H})}, v_{ij}^{(\boldsymbol{P})}$ 及び $\boldsymbol{R}_j^{(\boldsymbol{H})}, \boldsymbol{R}_j^{(\boldsymbol{P})}$ に事前分布を導入し，EM アルゴリズム [49] を用いた最大事後確率推定で各パラメータを求める．これにより，調波成分及び打撃成分のスペクトル平滑化と音源毎の空間伝搬系のパラメータ推定を行い，それらに基づいて分離を行う．

2.6　TFMBSS

TFMBSS [22] とは，時間周波数マスクで表現される音源モデルに基づく優決定条件 BSS である．この手法を解説する上で必要な理論及び TFMBSS の概要を 2.6.1–2.6.5 節で説明する．

2.6.1 近接作用素

TFMBSS を理解する上で基礎的な理論となる近接作用素（proximity operator）[20] を説明する．近接作用素は，ある下半連続な真凸関数 $f(\mathbf{x})$ に対して次式で定義される．

$$\mathrm{prox}_{\gamma f}(\mathbf{x}) = \underset{\mathbf{y}\in\mathbb{R}^N}{\mathrm{argmin}}\ f(\mathbf{y}) + \frac{1}{2\gamma}\|\mathbf{x}-\mathbf{y}\|^2 \tag{2.22}$$

ここで，$\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ は任意の N 次元実数ベクトルであり，$\gamma > 0$ は任意の重み係数である．$\|\cdot\|$ は ℓ_2 ノルムを表す．この作用素を分かりやすく説明するため，$N = 1$ 及び $f(\mathrm{x}) = |\mathrm{x}|$ である場合について考える．この時，入力を $\mathrm{x} = 2$ とすると，式 (2.22) は下式で書ける．

$$\mathrm{prox}_{\gamma|\cdot|}(2) = \underset{\mathrm{y}}{\mathrm{argmin}}\ |\mathrm{y}| + \frac{1}{2\gamma}(2-\mathrm{y})^2 \tag{2.23}$$

式 (2.23) における第二項は，任意の y と入力値 $\mathrm{x} = 2$ の二乗誤差である．この二乗誤差と入力関数 f（式 (2.23) においては絶対値関数）との和が最も小さくなる点を返すのが近接作用素である．これは，二乗誤差という罰則条件与えられた上で，入力関数 f の最小点を返すことに相当する．言い換えると，この操作 $\mathrm{prox}_{\gamma f}(\mathrm{x})$ は，入力値 x の近傍で，入力関数 f が小さくなる点を返す作用素である．

2.6.2 時間周波数マスク

次に，時間周波数マスクについて説明する．時間周波数マスクとは，観測信号のスペクトログラムのある時間周波数要素に対して，目的の分離信号の成分が存在しているかどうかを表す二次元行列である．ソフトマスクであれば，時間周波数マスクは 0 から 1 までの値で構成され，バイナリマスクであれば，0 又は 1 の値で構成される．この概要を Fig. 2.6 に示す．図のような赤，緑，青の音源から成る混合信号から赤の音源成分のみを取り出したい状況を仮定する．この時，赤の音源の部分を 1 それ以外を 0 とするようなバイナリの時間周波数マスクを構成し，混合信号のスペクトログラムと時間周波数マスクを要素毎に乗算することで，赤の音源とその他の音源を分離することが可能である．

2.6.3 主双対近接分離法

2.6.1 節の近接作用素を応用した，汎用的な最適化アルゴリズムである主双対近接分離法（primal-dual splitting method）[16]–[19] を解説する．主双対近接分離法では，次式の最小化問題を考える．

$$\min_{\mathbf{x}\in\mathbb{R}^N}\ f(\mathbf{x}) + g(\mathbf{x}) \tag{2.24}$$

ここで，$g(\mathbf{x})$ は下半連続な真凸関数であり非可微分関数とする．この最小化問題は，$g(\mathbf{x})$ が微分不可能であるため通常の最急降下法などでは，解くことができない．そこで主双対近接分

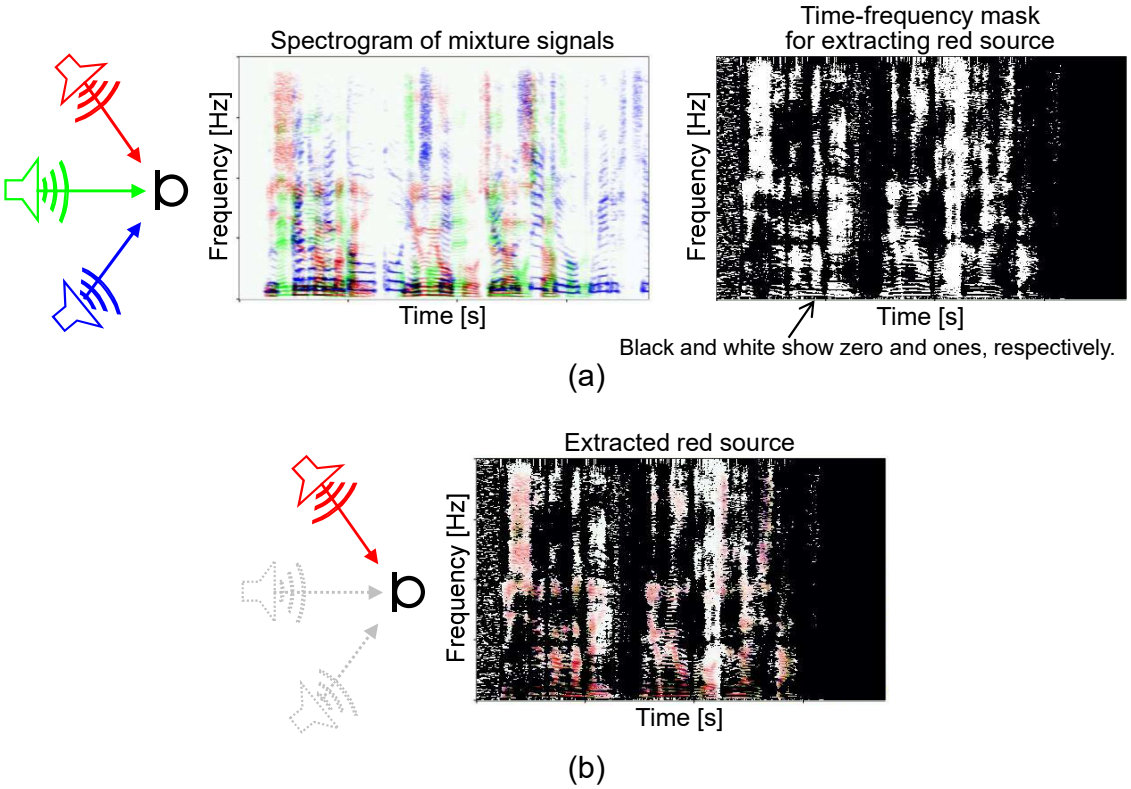

Fig. 2.6. Overview of time-frequency mask: (a) mixture signal and time-frequency binary mask for red source and (b) extracted red source by time-frequency binary mask.

離法では，近接作用素を用いて，$\mathbf{x} \in \mathbb{R}^N$ に対し以下の更新を与えることでこの問題を解く．

$$\mathbf{x}^{(n+1)} = \mathrm{prox}_{\gamma g}\left(\mathbf{x}^{(n)} - \gamma \nabla f(\mathbf{x}^{(n)})\right) \tag{2.25}$$

上式は，$f(\mathbf{x})$ に対する一般的な最急降下法の更新式を $g(\mathbf{x})$ に対する近接作用素の入力として与えることに相当する．この更新を，式 (2.22) のように展開すると次式で書ける．

$$\mathbf{x}^{(n+1)} = \underset{\mathbf{y}\in\mathbb{R}^N}{\mathrm{argmin}}\ g\left(\mathbf{x}^{(n)} - \gamma \nabla f(\mathbf{x}^{(n)})\right) + \frac{1}{2\gamma}\left\|\mathbf{y} - \left(\mathbf{x}^{(n)} - \gamma \nabla f(\mathbf{x}^{(n)})\right)\right\|^2 \tag{2.26}$$

よって，主双対近接分離法は，最急降下法によって $f(\mathbf{x})$ を小さくする点に飛ばし，その近傍で $g(\mathbf{x})$ も小さくなる点へと $\mathbf{x}$ を更新している．

上記の主双対近接分離法は，現実的な時間の中で近接作用素が計算可能であることを前提としており，この前提を prox 可能（proximable）と呼ぶ．即ち，式 (2.24) の問題において近接作用素の対象となる $g(\mathbf{x})$ が prox 可能であることが重要となる．そのため，主双対近接分離法を活用するには，解くべき問題を，prox 可能な $g(\mathbf{x})$ を含んだ式 (2.24) の形へと上手くモデル化する必要がある．音源分離の研究分野では，prox 可能な関数で表される音源モデルが多く提案されているため，主双対近接分離を応用が可能である．

Algorithm 1 primal-dual splitting BSS

Input: X, $\mathbf{w}^{[1]}$, $\mathbf{y}^{[1]}$, μ_1, μ_2, α

Output: $\mathbf{w}^{[k+1]}$

1: **for** $k = 1, 2, \cdots, K$ **do**
2: $\quad \widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1 \mathcal{I}} \left[\mathbf{w}^{[k]} - \mu_1 \mu_2 X^{\mathrm{H}} \mathbf{y}^{[k]} \right]$
3: $\quad \mathbf{z} = \mathbf{y}^{[k]} + X(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]})$
4: $\quad \widetilde{\mathbf{y}} = \mathbf{z} - \mathrm{prox}_{\frac{1}{\mu_2} \mathcal{P}}[\mathbf{z}]$
5: $\quad \mathbf{y}^{[k+1]} = \alpha \widetilde{\mathbf{y}} + (1 - \alpha) \mathbf{y}^{[k]}$
6: $\quad \mathbf{w}^{[k+1]} = \alpha \widetilde{\mathbf{w}} + (1 - \alpha) \mathbf{w}^{[k]}$
7: **end for**

2.6.4 主双対近接分離法を応用した線形分離アルゴリズム

音源分離の研究分野において，より良い音源モデルを目指した新しい BSS が提案される際，それらの音源モデルを統一的に扱えるフレームワークが存在するなら，既存の音源モデルとの比較が容易になる．この目的に対し，前述の主双対近接分離法を用いた幅広い音源モデルを統一的に扱える BSS アルゴリズムが提案されている [21]．この BSS アルゴリズムを Algorithm 1 に示す．ここで，X は多チャネル観測信号の複素スペクトログラム（$\boldsymbol{X}_1, \cdots, \boldsymbol{X}_M$）から構成される行列，$\mathbf{w}$ は全周波数の分離行列（$\boldsymbol{W}_1, \cdots, \boldsymbol{W}_I$）をベクトル化した変数である．$\mathcal{P}[\mathbf{z}]$ は音源モデルに対応する関数である（詳細な定義は文献 [21, 22] 参照）．このアルゴリズムは，前節で解説した主双対近接分離法と同じく，prox 可能な関数仮定されている場合にしか適用できない．即ち，関数として音源モデルを明示的に記述できる BSS (ICA, IVA, 及び ILRMA など）には適用可能であるが [21]，例えば，観測データを基に構成されるような，関数としての表現が未知の BSS には適用することができない．

2.6.5 TFMBSS の概要

前項に記載した問題を解決するために，関数形が未知の幅広い音源モデルを統一的に扱えるアルゴリズムとして，TFMBSS が提案されている [22]．TFMBSS のアルゴリズムを Algorithm 2 に示す．ここで，$\odot$ は要素毎の積を表わし，`generateMask`$(\cdot)$ はマスク生成関数である．Algorithm 2 は，Algorithm 1 の 4 行目のみが変更されたアルゴリズムである．Algorithm 1 における近接作用素の計算は，0 から 1 の範囲における閾値処理であり，これは時間周波数マスクを要素毎に適用する操作と等価関係にあることが解析されている [22]．これにより，明示的に音源モデルに対応する関数 $\mathcal{P}[\mathbf{z}]$ を書き下せなければならないという要件を回避した上で，上記のアルゴリズムと同様の働きをするアルゴリズムが実現されている．TFMBSS では，中間変数 $\mathbf{z}$ を引数とし分離をさらに促進するような時間周波数マスクを返す関数 `generateMask`$(\cdot)$ から生成される時間周波数マスク $\mathcal{M}$ を音源モデルとして活用する．

Algorithm 2 TFMBSS

Input: X, $\mathbf{w}^{[1]}$, $\mathbf{y}^{[1]}$, μ_1, μ_2, α

Output: $\mathbf{w}^{[K+1]}$

1: **for** $k = 1, 2, \cdots, K$ **do**
2: $\quad \widetilde{\mathbf{w}} = \mathrm{prox}_{\mu_1 \mathcal{I}}\left[\mathbf{w}^{[k]} - \mu_1 \mu_2 X^{\mathrm{H}} \mathbf{y}^{[k]}\right]$
3: $\quad \mathbf{z} = \mathbf{y}^{[k]} + X(2\widetilde{\mathbf{w}} - \mathbf{w}^{[k]})$
4: $\quad \mathcal{M} = \texttt{generateMask}(\mathbf{z})$
5: $\quad \widetilde{\mathbf{y}} = \mathbf{z} - \mathcal{M} \odot \mathbf{z}$
6: $\quad \mathbf{y}^{[k+1]} = \alpha \widetilde{\mathbf{y}} + (1 - \alpha)\mathbf{y}^{[k]}$
7: $\quad \mathbf{w}^{[k+1]} = \alpha \widetilde{\mathbf{w}} + (1 - \alpha)\mathbf{w}^{[k]}$
8: **end for**

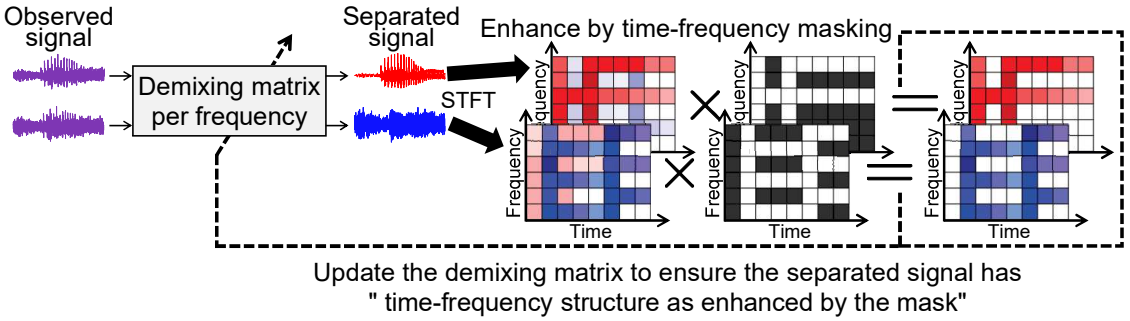

Fig. 2.7. Separation algorithm of TFMBSS.

従って，マスク生成関数 $\texttt{generateMask}(\cdot)$ を自由に入れ替えることで，様々な音源モデルを導入した BSS が実装できる．TFMBSS の分離アルゴリズムを Fig. 2.7 に示す．TFMBSS では，分離信号が時間周波数マスク $\mathcal{M}$ で強調されるような時間周波数構造を持つように分離行列 $\mathbf{w}$ を更新している．そのため，時間周波数マスク $\mathcal{M}$ の仮定する音源モデルに沿った分離が可能である．

2.7　本章のまとめ

本章では，本論文で重要な基礎理論について解説し，単一チャネル HPSS の非線形性に関する問題点及び TFMBSS のモデル構築における柔軟性を述べた．次章では，本章で解説した TFMBSS の利点を活用して，単一チャネル HPSS の問題点を解決するアルゴリズムを提案する．

第 3 章

調波打撃音モデルに基づく線形多チャネル音源分離

3.1　まえがき

前章では，従来の音源分離手法について詳しく解説し，単一チャネル HPSS の非線形性に対する問題点を述べた．本章では，時間周波数マスクを音源モデルとして用いる TFMBSS の柔軟性を活用し，単一チャネル HPSS と優決定 BSS の統合アルゴリズムを提案する．3.2 節では，単一チャネル HPSS の問題点及び優決定 BSS の利点を再度掘り下げ，本論文における提案アルゴリズムの動機について述べる．3.3 節では，提案アルゴリズムの概要を各ステップごとに詳しく解説し，3.4 節では，提案アルゴリズムの不安定性を解決するために，必要不可欠なスムージング法を提案し解説する．3.5 節では，本章のまとめを述べる．

3.2　提案アルゴリズムの動機

今日に至るまで，2.5 節のような HPSS は，音楽信号解析やリミックスの前処理として頻繁に使用されてきた．しかし，これらの分離手法は単一チャネル BSS であるため，調波音と打撃音を強力に分離することができる反面，非線形な音源分離であることに起因する音質の劣化が問題となる．例えば，音源分離の誤差成分が局所的に残留することによりミュージカルノイズ等の人工的な歪みが発生する場合がある．このような歪みは，後続するシステムの精度を低下させ，音楽信号の芸術的価値を損なわせるという問題に繋がる．この問題は，音楽信号のように芸術的価値が重要な信号においては深刻である．ここで，Fig. 3.1 に完全に分離された音源，単一チャネル OHPSS によって分離された音源，及び単一チャネル MHPSS によって分離された音源のスペクトログラムを比較した画像を示す．Fig. 3.1（a）及び（b）のスペクトログラムには，人工歪みが発生していることが分かる．一方，IVA や ILRMA のように観測信号が優決定条件（$M \geq N$）である場合は，線形な空間分離フィルタ（分離行列 $\boldsymbol{W}_i$）を推定することで，歪みの少ない自然な音源分離が可能となる．そのため，HPSS が仮定するような音

源モデルの分離を，優決定条件の下で実現することには有用性があると言える．

そこで本論文では，単一チャネル HPSS と優決定 BSS の統合アルゴリズムを提案する．今回提案するアルゴリズムにおいても，IVA や ILRMA と同じく優決定条件下で空間分離フィルタを推定する線形分離であるため，分離音源の歪みを最小限に抑えることができる．提案アルゴリズムは，TFMBSS の時間周波数マスク関数（Algorithm 2 におけるマスク関数 $\mathcal{M}$）に，従来の単一チャネル HPSS から生成した時間周波数マスクを音源モデルとして用いる．また TFMBSS の各反復時，時間周波数マスク $\mathcal{M}$ は HPSS によって更新されるため，反復推定が進むに応じて音源モデルの精度を上昇させる．さらに，この逐次的なマスク生成によるアルゴリズムの不安定さを軽減するため，スムージング法を提案する．このスムージング法により，マスク更新を安定化でき，分離音源の推定精度を向上させる．以降，単一チャネル HPSS である MHPSS 又は OHPSS のいずれかを選択して時間周波数マスクを生成し，それぞれ TFMBSS に活用したアルゴリズムの詳細を述べる．

3.3　提案アルゴリズムの概要

提案するアルゴリズムのブロック図を Fig. 3.2 に示す．まず，観測信号は STFT でスペクトログラムに変換する．次に，TFMBSS（Algorithm 2）によって，IVA や ILRMA と同様に線形な空間分離フィルタ（分離行列 $\boldsymbol{W}_i$）を推定する．その後，逆 STFT（inverse STFT: ISTFT）によって，分離された時間信号を得る．提案アルゴリズムでは，Algorithm 2 の 4 行目で，マスク関数 $\mathcal{M}$ として HPSS を用いる．HPSS を用いる際，2.5 節の OHPSS 又は MHPSS のいずれかを選択して適用するため，提案アルゴリズムは，選択する HPSS に応じて２つのバリエーションを持つことになる．

3.3.1　前後処理

TFMBSS の各反復において，入力 $\mathbf{z}$ を HPSS の入力に適した形にするため，`generateMask`$(\cdot)$ 内部で次の処理を施す．入力となる中間変数 $\mathbf{z} \in \mathbb{C}^{2IJ}$ は $\boldsymbol{Z}_H \in \mathbb{C}^{I\times J}$ と $\boldsymbol{Z}_P \in \mathbb{C}^{I\times J}$ に分割され，$\boldsymbol{Z}_H$ が調波成分，$\boldsymbol{Z}_P$ が打撃成分に相当する．この分割は単に変数の再定義であり，何の数値的処理にも該当しない．各反復において，$\mathbf{z}$ の半分は常に $\boldsymbol{Z}_H$ に割り当てられ、残りの半分は $\boldsymbol{Z}_P$ に割り当てられる．

提案アルゴリズムでは，TFMBSS の各反復における $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ のスケールの推定ができない．これは，IVA と ILRMA と同様に，分離行列 $\boldsymbol{W}_i$ がスケール不定であり，反復推定途中の分離行列によって得られる $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ の周波数 i 毎のスケールが適当なスケールとなってしまうことに起因する．この問題は，$\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ を入力として用いる HPSS の分離精度に影響を与えるため，スケールの固定は必要不可欠である．その解決として，プロジェクションバック法 [50, 51] を適用し，周波数毎の $\boldsymbol{Z}_H$ 及び $\boldsymbol{Z}_P$ のスケールを固定する．さらに，このスケール不定問題は，分離行列 $\boldsymbol{W}_i$ の推定後に式 (2.9) で得られる分離信号 $\boldsymbol{y}_{ij}$ に対しても同様である．分離信号 $\boldsymbol{y}_{ij}$ も，周波数 i 毎に不揃いなスケールで得られるため，そのまま ISTFT

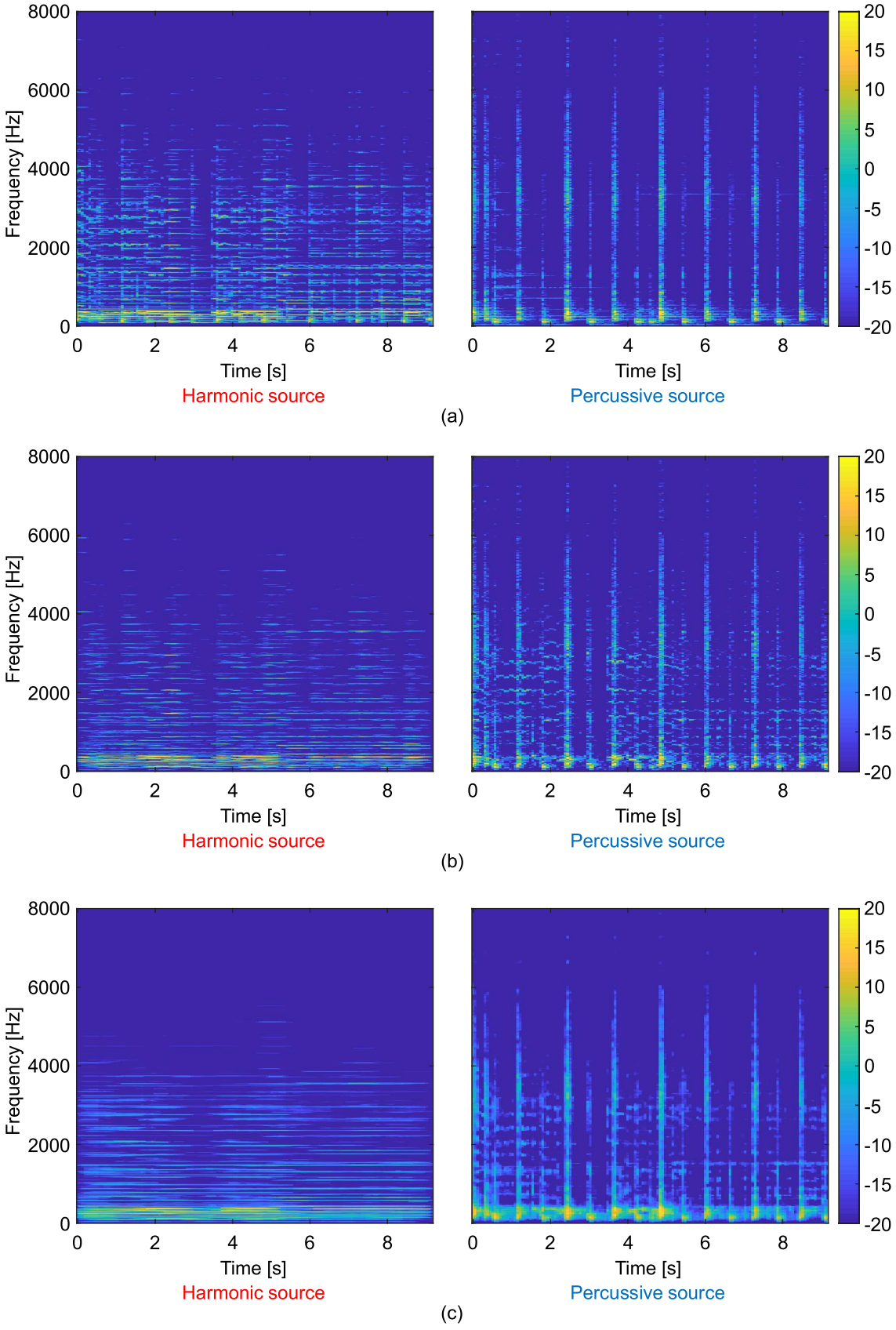

Fig. 3.1. Spectrograms of (a) oracle harmonic and percussive sources, (b) estimated harmonic and percussive sources obtained by OHPSS, and (c) estimated harmonic and percussive sources obtained by OHPSS.

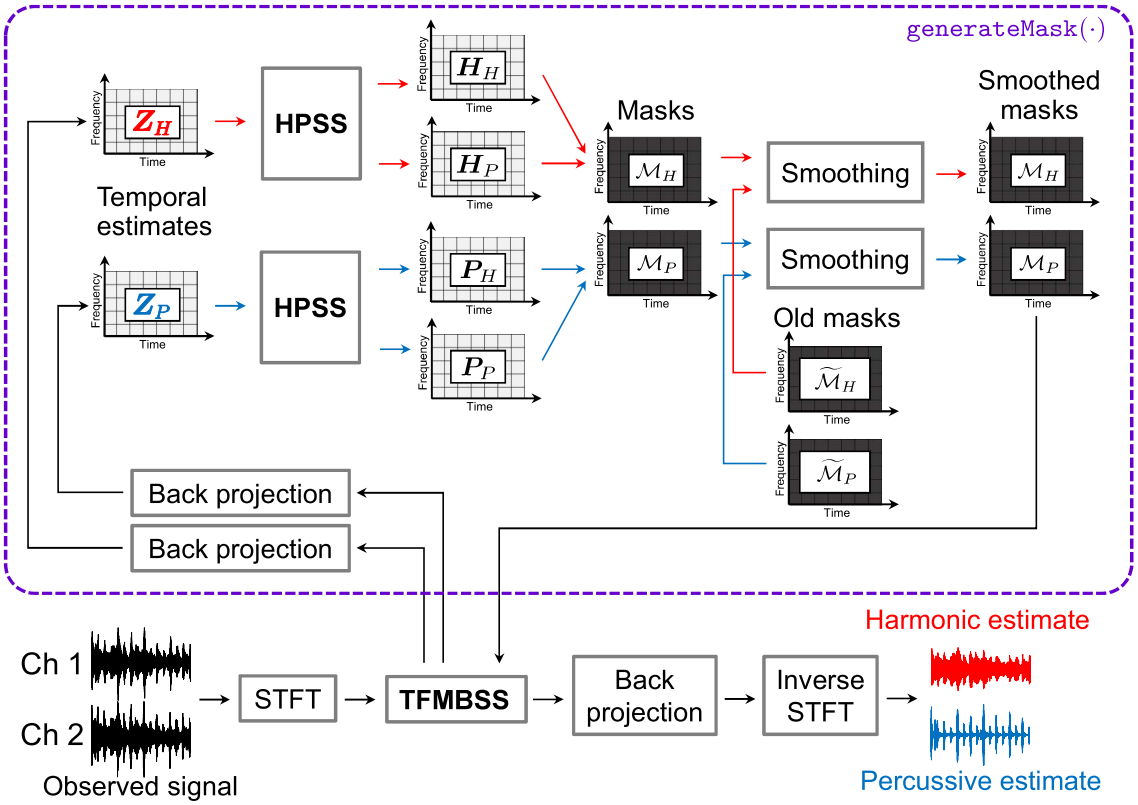

Fig. 3.2. Block diagram of proposed method.

を適用して時間信号に変換した場合，正しい分離信号は得られない．この解決策としても，同様にプロジェクションバック法を導入する．本論文では，プロジェクションバック法について分離信号 $\boldsymbol{y}_{ij}$ を用いて説明する．ここで，$\boldsymbol{y}'_{ijn}$ を次式のように定義する．

$$\boldsymbol{y}'_{ijn} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ y_{ijn} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.1}$$

即ち，$\boldsymbol{y}'_{ijn}$ は $\boldsymbol{y}_{ij}$ の n 番目の要素のみを残し，他を 0 とした分離信号である．プロジェクションバック法では，式 (3.1) の分離信号に対して，推定済の分離行列 $\boldsymbol{W}_i$ の逆行列 $\boldsymbol{W}_i^{-1}$ を次式

のように適用する．

$$\begin{bmatrix} \hat{y}_{ijn1} \\ \vdots \\ \hat{y}_{ijnm} \\ \vdots \\ \hat{y}_{ijnM} \end{bmatrix} = \boldsymbol{W}_i^{-1} \boldsymbol{y}'_{ijn} \tag{3.2}$$

ここで，得られた信号 $\hat{y}_{ijnm}$ は m 番目のマイクロホンで観測されたスケールに合わせた n 番目の分離信号である．式 (3.2) で表されるプロジェクションバック法により，スケールを全周波数で合わせた分離信号 $\hat{y}_{ijnm}$ が得られる．特に本論文でのプロジェクションバック法は，常に 1 番目のマイクロホンで観測されたスケールに合わせた信号を出力することとする．即ち，式 (3.2) の $\hat{y}_{ijn1}$ のみをプロジェクションバック法の出力として用いる．以降の項では，入力 $\mathbf{z}$ から出力 $\mathcal{M}$ までの処理を順を追って述べる．

3.3.2　OHPSS モデルにおける分離推定

前項の前処理を施した中間変数 $\boldsymbol{Z}_H$ と $\boldsymbol{Z}_P$ それぞれに対して，Fig. 3.2 に示すように，個別に HPSS を適用する．適用する HPSS は，OHPSS と MHPSS のいずれかから選択して適用する．OHPSS を適用する場合，混合信号（$|\boldsymbol{B}| \in \mathbb{R}_{\geq 0}^{I \times J}$）と調波信号及び打撃信号（$|\boldsymbol{H}|, |\boldsymbol{P}| \in \mathbb{R}_{\geq 0}^{I \times J}$）は，$\boldsymbol{Z}_H$ 又は $\boldsymbol{Z}_P$ のいずれかを用いて初期化される．$\boldsymbol{Z}_H$ の場合，以下のように初期化される．

$$|\boldsymbol{B}|^{\xi} = |\boldsymbol{Z}_H|^{\xi} \tag{3.3}$$

$$|\boldsymbol{H}|^{\xi} = \frac{1}{2} |\boldsymbol{Z}_H|^{\xi} \tag{3.4}$$

$$|\boldsymbol{P}|^{\xi} = \frac{1}{2} |\boldsymbol{Z}_H|^{\xi} \tag{3.5}$$

その後，OHPSS の反復推定アルゴリズムに従い，式 (2.14)–(2.15) の更新式を $\boldsymbol{Z}_H$ に繰り返し適用することで，$\boldsymbol{Z}_H$ の打撃成分及び調波成分が得られる．得られた信号を，添え字 $(\cdot)_H$ を用いて，$\boldsymbol{H}_H$ 及び $\boldsymbol{P}_H$ とする．同様の方法で，$\boldsymbol{Z}_P$ の場合も以下のように初期化される．

$$|\boldsymbol{B}|^{\xi} = |\boldsymbol{Z}_P|^{\xi} \tag{3.6}$$

$$|\boldsymbol{H}|^{\xi} = \frac{1}{2} |\boldsymbol{Z}_P|^{\xi} \tag{3.7}$$

$$|\boldsymbol{P}|^{\xi} = \frac{1}{2} |\boldsymbol{Z}_P|^{\xi} \tag{3.8}$$

そして，式 (2.14)–(2.15) の更新式を $\boldsymbol{Z}_P$ に繰り返し適用することで，$\boldsymbol{Z}_P$ の打撃成分及び調波成分が得られる．同じく，添え字 $(\cdot)_P$ を用いて，$\boldsymbol{H}_P$ 及び $\boldsymbol{P}_P$ とする．従って，OHPSS を 2 回実行することにより，2 組の振幅スペクトログラム（$|\boldsymbol{H}_H|, |\boldsymbol{P}_H|$）及び（$|\boldsymbol{H}_P|, |\boldsymbol{P}_P|$）が得られる．

3.3.3　MHPSS モデルにおける分離推定

MHPSS を適用する場合，観測信号 $\boldsymbol{B}$ を以下のように初期化する．

$$|\boldsymbol{B}| = |\boldsymbol{Z}_H| \tag{3.9}$$

その後，$\boldsymbol{Z}_H$ に対して式 (2.18) の処理を適用することで，調波成分 $\boldsymbol{H}_H$ 及び打撃成分 $\boldsymbol{P}_H$ を得る．同様に $\boldsymbol{Z}_P$ を用いて，以下のような初期化を行う．

$$|\boldsymbol{B}| = |\boldsymbol{Z}_P| \tag{3.10}$$

その後，$\boldsymbol{Z}_H$ に対して式 (2.19) の処理を適用することで，調波成分 $\boldsymbol{H}_P$ 及び打撃成分 $\boldsymbol{P}_P$ を得る．従って，MHPSS を 2 回実行することにより，2 組の振幅スペクトログラム（$|\boldsymbol{H}_H|, |\boldsymbol{P}_H|$）及び（$|\boldsymbol{H}_P|, |\boldsymbol{P}_P|$）が得られる．

3.3.4　時間周波数マスクの生成

3.3.2 節または 3.3.3 節で得られた信号（$|\boldsymbol{H}_H|, |\boldsymbol{P}_H|$）及び（$|\boldsymbol{H}_P|, |\boldsymbol{P}_P|$）は，`generateMask`$(\cdot)$ の出力が区間 $[0,\ 1]$ の値から成る時間周波数マスクでなければならない．本論文では，HPSS の結果に基づいて，以下のようなマスク $\mathcal{M}_H$ 及び $\mathcal{M}_P$ [23] を構築する．

$$\mathcal{M}_H = \frac{|\boldsymbol{H}_H|^2}{|\boldsymbol{H}_H|^2 + |\boldsymbol{P}_H|^2} \tag{3.11}$$

$$\mathcal{M}_P = \frac{|\boldsymbol{P}_P|^2}{|\boldsymbol{H}_P|^2 + |\boldsymbol{P}_P|^2} \tag{3.12}$$

ここで，式 (3.11) 及び式 (3.12) における演算は全ての要素毎に行われる．これらのマスクは，他の成分を排除して，調波成分や打撃成分を強調するものである．そのため，$\mathbf{z}$ には，どの成分を削減すべきかという情報が $\mathcal{M}$ によって与えられる．TFMBSS はこの情報に従って，分離フィルタを更新する．この時，Algorithm 2 の 5 行目のように，これらのマスクを $\mathbf{z} \in \mathbb{C}^{2IJ}$ に適用するため，マスク $\mathcal{M}_H \in [0,1]^{I\times J}$ 及び $\mathcal{M}_P \in [0,1]^{I\times J}$ を連結してベクトル化し，$\mathcal{M} \in [0,1]^{2IJ}$ を生成する．そして，`generateMask`$(\cdot)$ から出力される．

これら 3.3.2–3.3.4 節のステップによって，$\mathcal{M}$ を構築する `generateMask`$(\cdot)$ が定義される．この時，生成された時間周波数マスク $\mathcal{M}$ の精度は HPSS が正しく分離できたかに依存しているが，HPSS では信号を正しく分離できない場合も存在する．この時間周波数マスク $\mathcal{M}$ の精度は TFMBSS の安定性に影響を与えるため，次節で説明するスムージング法（Fig. 3.2 の smoothing block）を提案する．

3.4 マスクのスムージング

TFMBSS におけるマスク生成関数は任意であり，その自由度から，上記の HPSS に基づくマスク生成関数の構築を実現している．しかし，どのようなマスク生成関数でも合理的な TFMBSS アルゴリズムが得られる訳ではなく，安定かつ有用であるマスク生成関数のみが，性能の良いアルゴリズムを与えることができる．特に，TFMBSS は主双対近接分離をベースにしているため，ある反復更新において，更新後の状態は更新前の状態から近傍にあることを前提としている．そのため，TFMBSS は，反復間の状態が大きく変化すると安定した音源分離ができない場合がある．提案アルゴリズムは，反復毎に HPSS でマスク $\mathcal{M}$ の再構築を行うことから，マスク $\mathcal{M}$ が大幅に変動する．よって，提案アルゴリズムは，上記の前提に反するため最適化としての安定性に欠けてしまう．

この問題に対処するために，提案アルゴリズムではマスクを生成する度に，1 反復前のマスクとの平滑化を行うスムージング法を提案する．この操作により，反復間のマスクの大きな変化を避けることで TFMBSS の最適化を安定させる．このマスクのスムージング処理は次式で表される．

$$\mathcal{M} = \mathcal{M}^{\beta} \odot \widetilde{\mathcal{M}}^{\beta_{\mathrm{old}}}, \tag{3.13}$$

ここで，$\widetilde{\mathcal{M}}$ は 1 反復前の時間周波数マスクである．加えて β 及び β_{old} はスムージングパラメータであり，$\beta + \beta_{\mathrm{old}} = 1$ とする．β_{old} を増加させることで，1 反復前の時間周波数マスク $\widetilde{\mathcal{M}}$ の影響が強くなるため，スムージングが強力に適用される．$\beta = 1$（$\beta_{\mathrm{old}} = 0$）の場合はスムージングを適用しないことと同義である．提案アルゴリズムの場合，上記のスムージング法は，以下のように $\mathcal{M}_H$ と $\mathcal{M}_P$ に独立に適用される．

$$\mathcal{M}_H = \mathcal{M}_H^{\beta} \odot \widetilde{\mathcal{M}}_H^{\beta_{\mathrm{old}}}, \tag{3.14}$$

$$\mathcal{M}_P = \mathcal{M}_P^{\beta} \odot \widetilde{\mathcal{M}}_P^{\beta_{\mathrm{old}}}, \tag{3.15}$$

これらのスムージング法は，式 (3.11) 及び式 (3.12) によるマスク計算の直後に実行される．安定性と性能はスムージングパラメータ β 及び β_{old} に依存するトレードオフの関係にあるため，パラメータの検討及び分離性能への影響については 4.5 節で議論する．

なお，このスムージング法の適用可能性は，本論文の提案アルゴリズムに限られるものではなく，TFMBSS に基づくあらゆる BSS アルゴリズムを安定化させることができる．従って，既に文献で提案されている TFMBSS を用いたアルゴリズム [22, 23] 及び今後提案される TFMBSS を用いたアルゴリズムに有用性を示すことが予想される．

3.5 本章のまとめ

本章では，従来の単一チャネル HPSS が持つ有用な音源モデルと，優決定 BSS である TFMBSS による線形分離を組み合わせたアルゴリズムを提案し，ステップ毎に詳細に解説し

た．さらに提案アルゴリズムにおいて，不安定性の問題に対する解決策としてスムージング処理を提案し解説した．次章では，この提案アルゴリズムのパラメータ検討及び有用性を確認するために複数の実験を行い考察する．

第 4 章

評価実験

4.1　まえがき

前章で提案したアルゴリズムの有用性を実験的に確認する．音楽信号を調波音と打撃音に分離するような音源分離実験を行い，性能評価を行った．4.2 節では，本実験における実験条件を詳細に説明する．4.3 節では，OHPSS の反復回数に対する分離性能の変化を確認し，4.4 節では，MHPSS のフィルタサイズに対する分離性能の変化を確認し，4.5 節では，スムージング法の有用性を実験的に示す．4.6 節では，MHPSS に基づく時間周波数マスクを用いた TFMBSS（MHPSS モデル）と，OHPSS に基づく時間周波数マスクを用いた TFMBSS（OHPSS モデル）の性能比較を行い，4.7 節では，既存の優決定 BSS や既存の HPSS と提案アルゴリズムの性能比較を行う．4.8 節では，本章のまとめを述べる．

4.2　実験条件

提案アルゴリズムの有効性を確認するために，音楽信号中のドラムとそれ以外の楽器音（後述のその他の音源（other）に該当）の音源分離実験を行った．本実験では，SiSEC2016 [52] の DSD100 データセットを使用した．DSD100 はトレーニングセット（Dev）とテストセット（Test）の 2 つのデータセットが存在し各 50 曲が収録されている．各曲は様々なスタイルの楽曲で構成されており，ボーカル音源（vocals），ベース音源（bass），ドラム音源（drums），及びその他の音源（other）が音源毎に収録されている．ここで，その他の音源（other）は，鍵盤楽器，管楽器，及び弦楽器などの音階を持った楽器の混合音であり，その編成は楽曲に依存する．DSD100 のテストセットの中でアルファベット順で並べた音源から 3〜6，8，11，13〜19，36 番目の楽曲 14 曲，DSD100 のトレーニングセットの中でアルファベット順で並べた音源から 1，3，9，15，23，27 番目の楽曲 6 曲のドラム音源（drums）とその他の音源（other）を選び，それぞれ Song ナンバー 1〜14 及び 15〜20 に割り当てた．選定の基準は，楽曲自体に打撃音と調波音がバランス良く存在しているか又はシンセドラムなどではなく一般的なドラム音かどうかである．これらのドライソースを，Fig. 4.1 に記載のマイクロホン間隔 5.66 cm

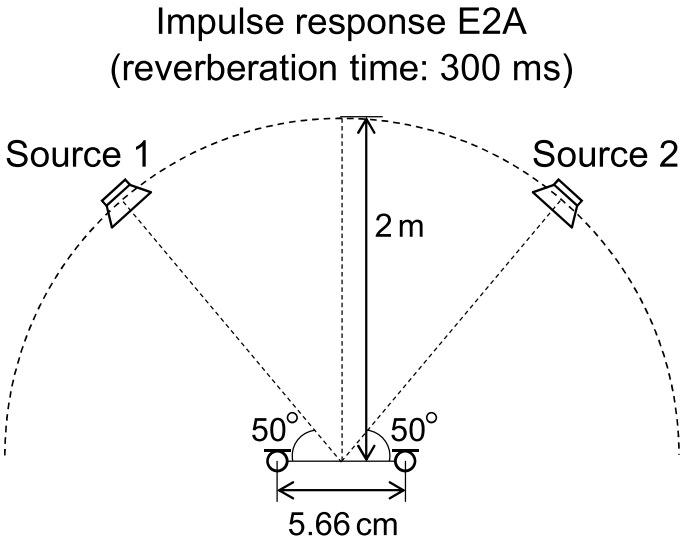

Fig. 4.1. Source and microphone arrangements to produce mixture signals used in experiment. These impulse responses are included as E2A in RWCP dataset.

Table 4.1. Experimental conditions

Window function in STFT	Hann window
Window length in STFT	128 ms
Shift length in STFT	64 ms
Parameters in OHPSS	$\kappa_H = 1.02$, $\kappa_P = 1.01$ $\xi = 1$, $\rho = 1/2$
Parameters in TFMBSS	$\alpha = 0.25$ $\mu_1 = \mu_2 = 1.0$
Number of iterations in TFMBSS	500

及びマイクロホンを中点とした半径 2m の円上で音源方位 50°&130° の RWCP データベース [53] 収録 E2A インパルス応答（残響長 300 ms）で畳み込み，多チャネル混合信号を作成した．その他の実験条件は Table 4.1 に示す．評価指標に信号対歪み比（source-to-distortion ratio: SDR）[54] の改善量を用いた．

4.3　OHPSS の反復回数における影響の検証

OHPSS を用いた提案アルゴリズムでは，TFMBSS を 1 回反復する毎に，任意の回数で OHPSS の更新式 (2.14)–(2.15) を反復計算している．そして，OHPSS の反復回数は提案アルゴリズムの分離性能に影響を与えることが実験的に分かっている．そこで，OHPSS の反復回数を変化させることによる SDR 改善量の変化を調査した．

提案アルゴリズムにおける全 20 曲の平均 SDR 改善量を Table 4.2 に示す．ここで，表にお

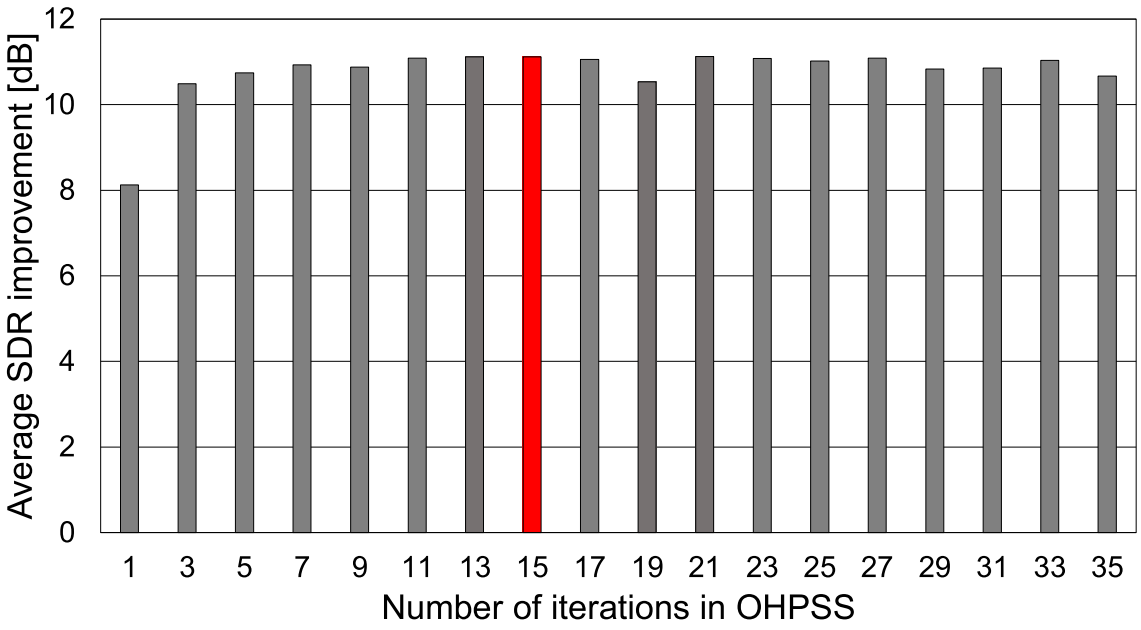

Fig. 4.2. Average SDR improvements of OHPSS-based proposed method with various numbers of iterations in OHPSS.

ける両手法のスムージングパラメータは $\beta_{\mathrm{old}} = 0.75$ 及び $\beta = 0.25$ であり，これはアルゴリズムの安定性と分離性能のバランスが良い設定値である（次節の実験結果に基づく）．OHPSS の反復更新式 (2.14)–(2.15) は、局所最小点である式 (2.11) に急速に収束するため，OHPSS の反復回数が 15 付近以上の場合，提案アルゴリズムの性能は飽和すると考えられる．以降の実験では，この結果に基づき，提案アルゴリズムにおける HPSS の反復回数を 15 回と設定する．

4.4　MHPSS のフィルタサイズにおける影響の検証

MHPSS を用いた提案アルゴリズムでは，メディアンフィルタのサイズ $2D + 1$ を設定してから推定を行う．このときのフィルタサイズを変化させることによる SDR 改善量の変化を調査した．

提案アルゴリズムにおける全 20 曲の平均 SDR 改善量を Fig. 4.3 に示す．ここで，提案アルゴリズムにおいて，各次元方向のフィルタの最適なサイズはほぼ等しいことが実験的に確認できている．そのため，時間方向と周波数方向のフィルタサイズは同一に固定した．前節と同じく，本実験におけるスムージングパラメータは $\beta_{\mathrm{old}} = 0.75$ 及び $\beta = 0.25$ である．Fig. 4.3 の結果より，本実験条件では，フィルタサイズは 19 点が最適であることが確認された．スムージングパラメータを変更した際，最適な数値が変動することも実験的に確認しているが，おおよそ 19 点周辺の値であった．以降の実験では，この結果に基づき，提案アルゴリズムにおけるフィルタリングサイズを 19 点と設定する．

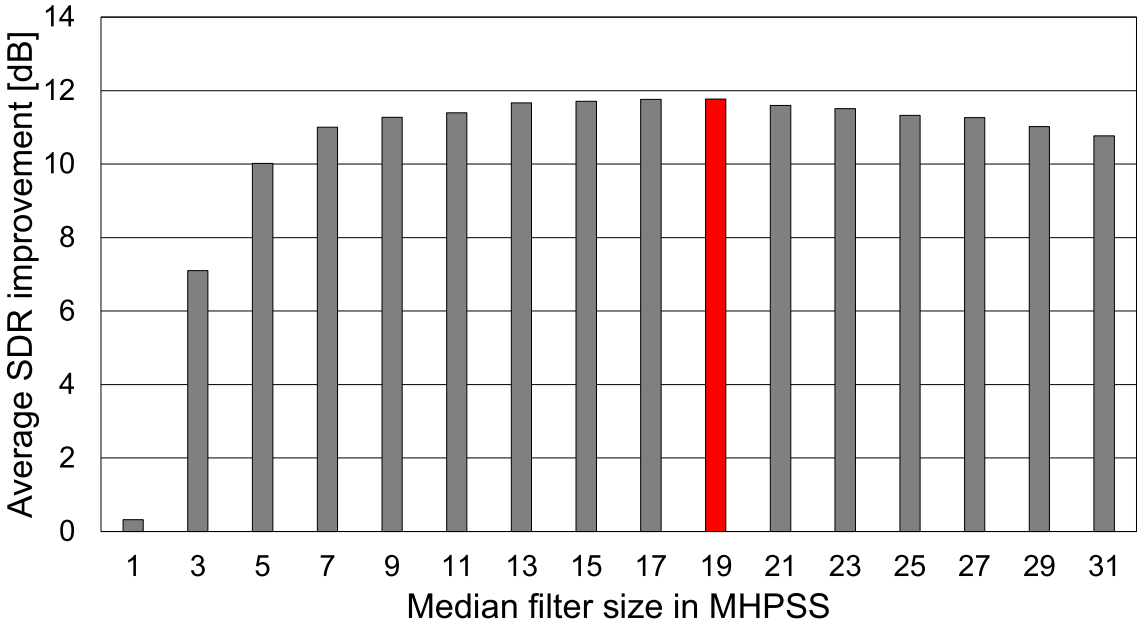

Fig. 4.3. Average SDR improvements of MHPSS-based proposed method with various filter sizes in MHPSS.

4.5　マスクのスムージングにおける影響の検証

本節では，3.4 節で提案したスムージング法の有用性を検証する．式 (3.15) の β_{old} 及び β のみを変えた場合の，MHPSS モデルにおける SDR 改善量の収束挙動の一例を Fig. 4.4 に示し，その他の楽曲における収束挙動は付録 A に示す．また，OHPSS モデルの収束挙動の一例を Fig. 4.5 に示し，その他の楽曲における収束挙動は付録 B に示す．ここで，$\beta = 1 - \beta_{\mathrm{old}}$ であり，$\beta_{\mathrm{old}} = 0$ がスムージング法を用いない場合のアルゴリズムに相当する．結果から，スムージング法によって，提案アルゴリズムの収束挙動を明らかに安定化できていることが確認できる．さらに，どちらの音源モデルに基づく提案アルゴリズムも安定化されているため，スムージング法は音源モデルによらず，安定性を発揮すると言える．

Figs. 4.4 及び 4.5 から確認できる通り，安定性と収束の速さ及び収束値はトレードオフの関係にある．従って，スムージング法は安定性と引き換えに，本来持つ潜在的な分離性能を低下させていると考えられる．提案アルゴリズムにおける全 20 曲の平均 SDR 改善量を，Tables 4.2 及び 4.3 に示す．これらの表より，スムージング法を強力に適用した場合，提案アルゴリズム性能が低下することがわかる．よって，提案アルゴリズムにおける β 及び β_{old} の条件は，収束値と安定性のトレードオフを考慮して，$\beta_{\mathrm{old}} = 0.75$ 及び $\beta = 0.25$ が最適と言える．これを踏まえ，以降全ての実験では $\beta_{\mathrm{old}} = 0.75$ 及び $\beta = 0.25$ を採用する．

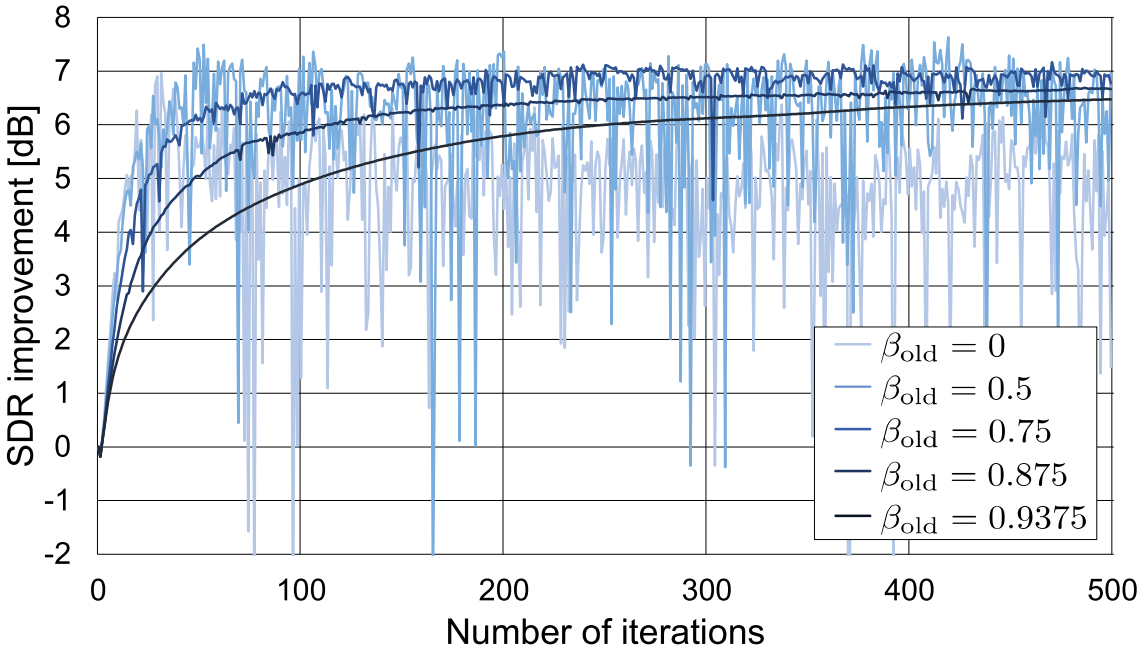

Fig. 4.4. Example of convergence behaviors of OHPSS-based proposed method with various β_{old} and β (song no. 1).

Table 4.2. Average SDR improvements of proposed method with various smoothing parameters β and β_{old}

β	β_{old}	Average SDR improvement [dB]
1	0	8.21
0.5	0.5	9.90
0.25	0.75	**11.12**
0.125	0.875	10.96
0.0625	0.9375	10.67

4.6　OHPSS モデルと MHPSS モデルの性能比較実験

本節では，OHPSS モデルと MHPSS モデルの性能比較を行う．Fig. 4.6 は，計算に要した経過時間に対する，データセット 20 曲全ての平均 SDR 改善量の比較である．ここで，図中の各マーカーは TFMBSS の各反復更新が終了した時点の経過時間である．即ち，各マーカーの間隔が 1 反復更新に要する時間に該当する．この結果より，OHPSS モデルは 1 反復更新での収束度は高いが，1 反復更新に要する時間は長いことが確認された．対して，MHPSS モデルは，1 反復更新での収束度は低いが，1 反復更新に要する時間は短い．よって状況に応じて，どちらのモデルも最適となり得る．これらの特徴は，MHPSS が単にフィルタを適用する手法

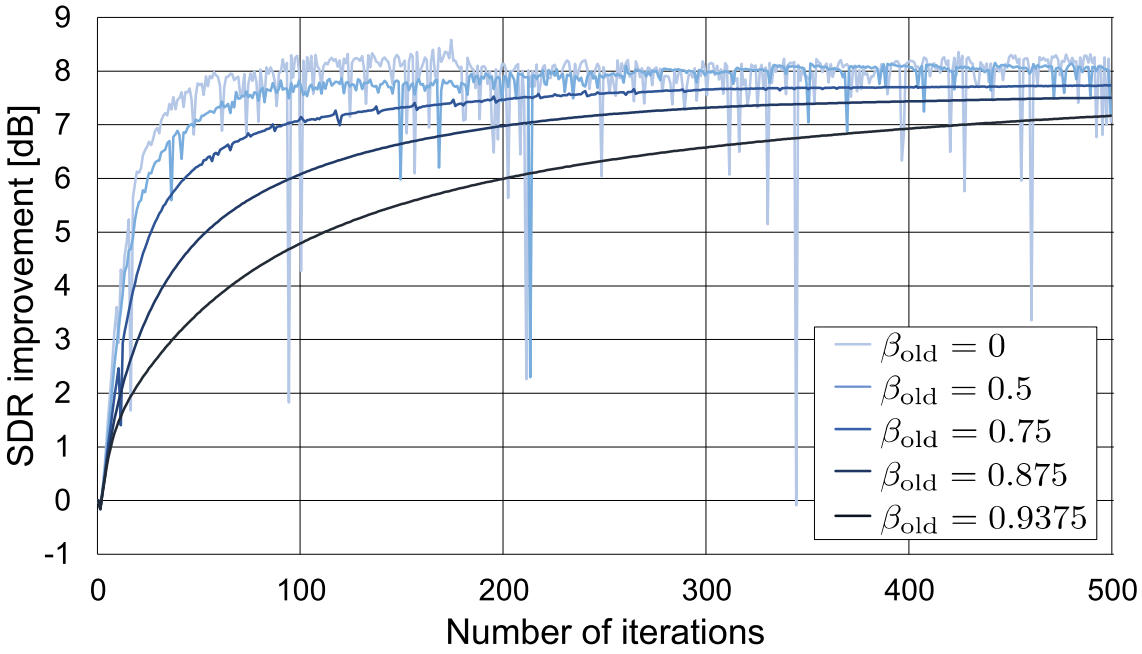

Fig. 4.5. Example of convergence behaviors of MHPSS-based proposed method with various β_{old} and β (song no. 1).

Table 4.3. Average SDR improvements of proposed method with various smoothing parameters β and β_{old}

β	β_{old}	Average SDR improvement [dB]
1	0	11.21
0.5	0.5	11.44
0.25	0.75	**11.77**
0.125	0.875	11.49
0.0625	0.9375	10.76

であるのに対し，OHPSS が反復推定手法であることに起因すると考察できる．平均的な性能については，MHPSS モデルの方が優れていた．

4.7　他の従来手法との性能比較実験

提案アルゴリズムの有効性を確認するために，単一チャネル OHPSS，単一チャネル MHPSS，多チャネル HPSS，補助関数法 IVA，ILRMA，OHPSS モデルに基づく TFMBSS，及び MHPSS モデルに基づく TFMBSS の 7 種類の手法で性能を比較した．ここで，多チャネル HPSS [46] では，Table 4.4 に記載の 2 種類の条件のパラメータ （params. in [46] 及び fine-tuned params.）を使用した．単一チャネル OHPSS，単一チャネル MHPSS，補助関数法 IVA，ILRMA，及び提案アルゴリズムでは，Hann window を用いており，窓長及びシフ

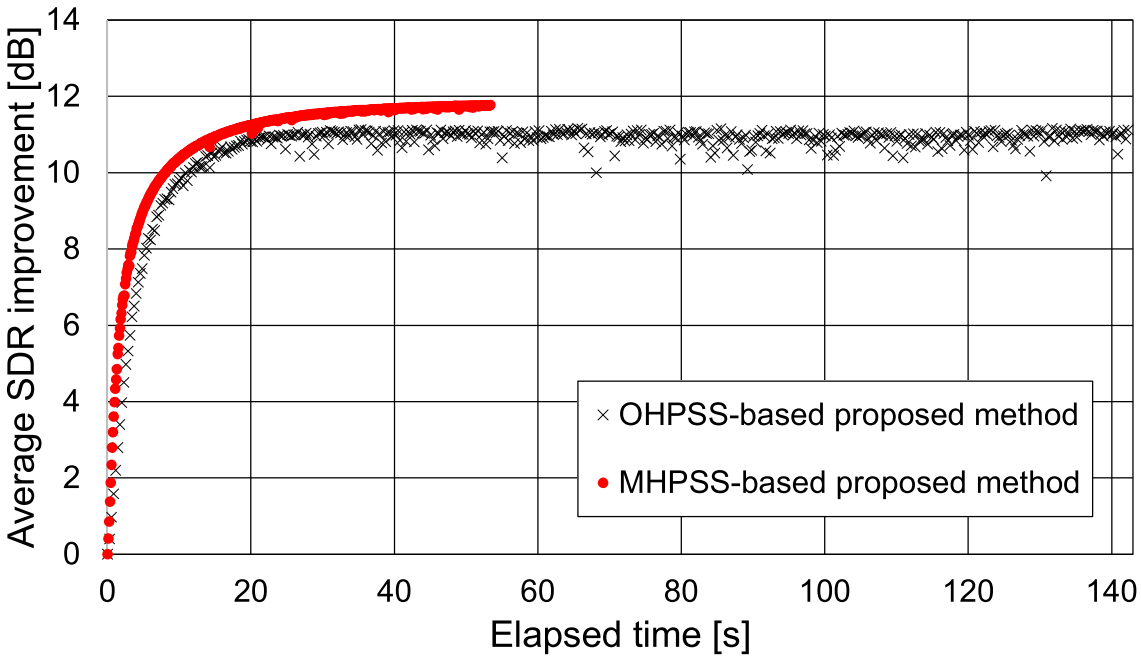

Fig. 4.6. Average convergence behaviors of SDR improvements for OHPSS-based and MHPSS-based proposed methods in terms of elapsed time.

ト長はそれぞれ 128ms，64ms である．その他の条件を Table 4.4 に記載する．

従来の BSS と性能比較における 20 曲全ての平均 SDR を Fig. 4.7 に示す．各楽曲における SDR 比較は付録 C に示す．さらに，20 曲全てにおける SDR 改善量のボックスプロットを Fig. 4.8 に示す．提案アルゴリズムは，ほぼ全ての楽曲において従来手法を上回っており，有用性が確認できる．ボックスプロットより，OHPSS モデルでは，分離性能の平均値は低く中央値は高いため，OHPSS が分離を得意とする曲は多いが平均的な分離性能は低いと考察できる．対して MHPSS モデルでは，中央値は低く平均値は高いため，MHPSS が分離を得意とする曲は少ないが平均的な分離性能は高いと考察できる．よって，どちらの音源モデルにも有用性があると言える．

4.8 本章のまとめ

本章では，提案アルゴリズムの各パラメータの検討及び既存手法との性能比較を行った．最適パラメータは分離する楽曲によって若干変化するが，それぞれ，OHPSS の反復回数は 15 回，MHPSS のフィルタサイズは 19，両モデル共にスムージングパラメータは $\beta_{\mathrm{old}} = 0.75$ 及び $\beta = 0.25$ となった．性能比較では，既存の単一チャネル HPSS 及び優決定 BSS に比べて性能を上回る結果となった．提案アルゴリズムは，調波音と打撃音の分離において十分な性能向上をもたらしたといえる．次章では，本論文における総括とした結論を述べる．

Table 4.4. Conditions for other HPSS and BSS algorithms

Parameters for multichannel HPSS described in [46]	128 ms Hann window with 1/2 shift $\alpha_h = \alpha_p = 10$, $m_h = m_p = 5$ $\gamma_1 = 0.5, \gamma_2 = 1$
Fine-tuned parameters for multichannel HPSS	16 ms Hann window with 1/2 shift $\alpha_h = \alpha_p = 5$, $m_h = m_p = 10$ $\gamma_1 = \gamma_2 = 1$
Number of bases in ILRMA	10 for each source
Number of iterations	20 for single-channel HPSS 15 for multichannel HPSS 30 for AuxIVA 100 for ILRMA

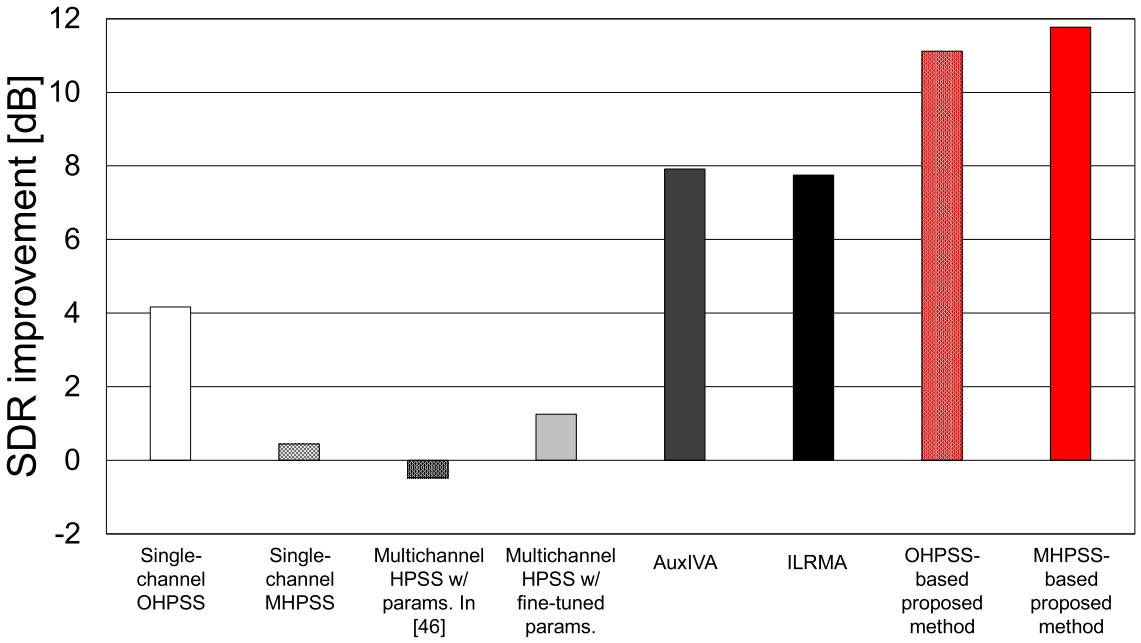

Fig. 4.7. SDR improvements of ILRMA, IVA, conventional HPSS, and pro-posed methods (Average of all songs).

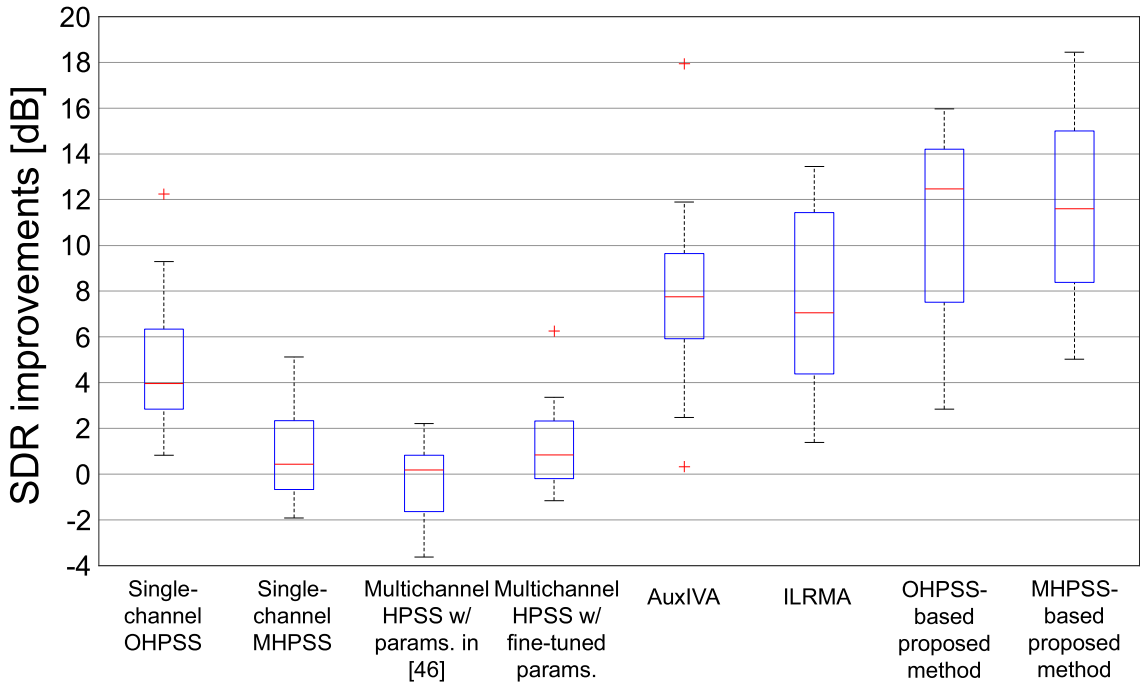

Fig. 4.8. Box plots of SDR improvements for 20 songs in each method. In each method, the blue box shows the range of 25–75 percentiles, red horizontal line in each box shows a median value, and the red cross mark shows outliers.

第 5 章

結言

本論文では，調波音と打撃音の線形分離を目的とし，調波打撃音モデルに基づく時間周波数マスクを TFMBSS に活用した音源分離アルゴリズムを新たに提案した．また，TFMBSS の最適化を安定させるために，時間周波数マスクのスムージング法も導入した．実験により，提案アルゴリズムにおける複数のパラメータに対して，最適なパラメータを探索した．さらに，提案アルゴリズムの不安定性に関する問題を，推定の収束挙動の安定と収束速度のトレードオフを考慮した適当な β_{old} 及び β によってスムージングすることで解決できることも実験的に示した．そして，提案アルゴリズムは，従来の HPSS 及び既存の優決定 BSS と比較して音質が向上したことを実験的に示し，提案アルゴリズムの有用性を明らかとした．

最後に今後の展望を述べる．今回は MHPSS モデル及び OHPSS モデルを用いた調波音と打撃音の分離を対象としたが，提案した TFMBSS による線形分離アルゴリズムやスムージング法は，他の音源分離手法にも適用が可能である．よって，さらに多くの種類の調波打撃音モデルを活用することで，より良い調波音と打撃音における音源モデルの探求が促進されることを期待する．加えて，分離音声信号分離やノイズ除去など，別の目的を持った音源分離手法にも適用が可能であるためそれらの分野への応用も期待できる．アルゴリズムの改良によっては，複数チャネルのそれぞれに別の音源モデルに基づくマスクを適用することも可能であると考えられるため，この提案アルゴリズムには柔軟な発展の余地が大きく残されている．

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講師に心より感謝申し上げます．北村大地講師には，本研究分野における基礎的な知識から研究に関する詳細な議論など，細部にわたるまで丁寧にご指導いただきました．この３年間という期間の中で，社会人としての一般的な礼儀作法，資料作成スキル，文章の推敲技術や日々のスケジュール管理方法など，挙げればきりがないほどのご指導を頂きました．今では，研究室に配属される自分とは別人のような成長を遂げたと実感しています．これまで本当にありがとうございました．

本論の副査である村上幸一准教授及び柿元健准教授には，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．早稲田大学の矢田部浩平講師には，共同研究を通じミーティング及び論文添削での細かな指摘や，多数の知識のご教授を頂きました．心より感謝申し上げます．

北村研究室同期の岩瀬佑太氏は，研究室のムードメーカーとして研究室を明るく元気づけ，研究しやすい環境作りをしてくださいました．彼の枠に囚われない考え方には，日ごろから多くのインスピレーションを頂き，自分自身も成長できたと感じています．北村研究室同期の梶谷奈未氏は，日頃の生活面において，窮地の際には手厚いサポートを頂き，いつも自分の心の支えとなってくださいました．彼女からの数々のご支援があったからこそここまで来れたと断言できます．北村研究室同期の渡辺瑠伊氏には，普段から専門的知識及び情報分野における一般的な知見の確立に多くのアドバイスを頂きました．更に，書面作成におけるまとまったデザイン性や研究におけるソフトウェア環境の参考として多くのヒントを日頃から頂きました．また，研究室の後輩には，普段から明るく接して頂き楽しく研究生活を送ることができました．加えて，後輩のモチベーションの高さには圧倒されることも多く，自分の日ごろの頑張りが本当に足りているのか見つめ直すという機会も頂きました．その他にも，元研究室メンバーの山地修平氏からは，研究発表の際に応援の言葉で背中を押していただくなど，多くの良きメンバーに恵まれたと実感しています．

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．自分自身の研究生活は，皆様全員の支えがあってこそだと実感しています．今まで自分を支えてくださった皆様，これまで本当にありがとうございました．

参考文献

[1] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[2] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, "A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF," *APSIPA Transactions on Signal and Information Processing*, vol. 8, no. e12, pp. 1–14, 2019.

[3] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[4] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1–3, pp. 21–34, 1998.

[5] H. Sawada, R. Mukai, S. Araki, and S. Makino, "Convolutive blind source separation for more than two sources in the frequency domain," *in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2004, vol. 3, pp. III-885–III-888.

[6] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 2, pp. 666–678, 2006.

[7] H. Sawada, R. Mukai, S Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530–538, 2004.

[8] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," *in Proc. International Conference on Independent Component Analysis and Signal Separation*, 2006, pp. 601–608.

[9] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 70–79, 2006.

[10] N. Ono, "Stable and fast update rules for independent vector analysis based on

auxiliary function technique," *in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2011, pp. 189–192.

[11] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.

[12] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation with independent low-rank matrix analysis," *In Audio Source Separation*, S. Makino, Ed., pp. 125–155, Springer, Cham, 2018.

[13] 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.

[14] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *in Proc. Neural Information Processing Systems*, 2000, pp. 556–562.

[15] C. Févotte, N. Bertin, J.-L. Durrieu, "Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis," *Neural Computation*, vol. 21, no. 3, pp. 793–830, 2009.

[16] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," *Fixed-point Algorithms for Inverse Problems in Science and Engineering*, pp. 185–212, Springer, 2011.

[17] N. Parikh and S. Boyd, "Proximal algorithms," *Foundations and Trends in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[18] N. Komodakis and J. C. Pesquet, "Playing with duality: An overview of recent primal-dual approaches for solving large scale optimization problems," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 31–54, 2015.

[19] M. Burger, A. Sawatzky, and G. Steidl, "First Order Algorithms in Variational Image Processing," *Splitting Methods in Communication, Imaging, Science, and Engineering*, pp. 345–407, Springer, 2016.

[20] 小野峻佑, "近接分離アルゴリズムとその応用," オペレーションズ・リサーチ = *Communications of the Operations Research Society of Japan:* 経営の科学, vol. 64, no. 6, pp. 316–325, 2019.

[21] K. Yatabe and D. Kitamura, "Determined blind source separation via proximal splitting algorithm," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018, pp. 776–780.

[22] K. Yatabe and D. Kitamura, "Time-frequency-masking-based determined BSS with application to sparse IVA," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 715–719.

[23] K. Yatabe and D. Kitamura, "Determined BSS based on time-frequency masking

and its application to harmonic vector analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1609–1625, 2021.

[24] A. R. López, N. Ono, U. Remes, K. Palomäki, and M. Kurimo, "Designing multichannel source separation based on single-channel source separation," *in Proc. International Conference on Acoustics, Speech, and Signal Processing*, 2015, pp. 469–473.

[25] N. Ono, K. Miyamoto, J. Le Roux, H. Kameoka, and S. Sagayama, "Separation of a monaural audio signal into harmonic/percussive components by complementary diffusion on spectrogram," *in Proc. 16th European Signal Processing Conference*, 2008, pp. 1–4.

[26] H. Tachibana, T. Ono, N. Ono, and S. Sagayama, "Melody line estimation in homophonic music audio signals based on temporal-variability of melodic source," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 425–428.

[27] H. Tachibana, H. Kameoka, N. Ono, and S. Sagayama, "Comparative evaluations of various harmonic/percussive sound separation algorithms based on anisotropic continuity of spectrogram," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 465–468.

[28] H. Tachibana, N. Ono, and S. Sagayama, "Singing voice enhancement in monaural music signals based on two-stage harmonic/percussive sound separation on multiple resolution spectrograms," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 228–237, 2013.

[29] H. Tachibana, N. Ono, H. Kameoka, and S. Sagayama, "Harmonic/percussive sound separation based on anisotropic smoothness of spectrograms," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 12, pp. 2059–2073, 2014.

[30] D. FitzGerald, "Harmonic/percussive separation using median filtering," *in Proc. Proceedings of the International Conference on Digital Audio Effects*, vol. 13, 2010.

[31] N. Ono, K. Miyamoto, H. Kameoka, L RouxJonathan, Y. Uchiyama, E. Tsunoo, T. Nishimoto, and S. Sagayama, "Harmonic and percussive sound separation and its application to MIR-related tasks," *Advances in Music Information Retrieval*, pp. 213–236, 2010.

[32] E. Tsunoo, N. Ono, and S. Sagayama, "Rhythm map: Extraction of unit rhythmic patterns and analysis of rhythmic structure from music acoustic signals," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 185–188.

[33] H. Tachibana, T. Ono, N. Ono, and S. Sagayama, "Melody line estimation in homophonic music audio signals based on temporal-variability of melodic source," *in*

Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 425–428, 2010.

[34] C. Hus, D. Wang, and J. Jang, "A trend estimation algorithm for singing pitch detection in musical recordings," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 393–396.

[35] J. Reed, Y. Ueda, M. Siniscalchi, Y. Uchiyama, S. Sagayama, and C. Lee, "Minimum Classification Error Training to Improve Isolated Chord Recognition," *in Proc. 10th International Society for Music Information Retrieval Conference*, 2009, pp. 609–614.

[36] J. Futrelle and J. S. Downie, "Interdisciplinary research issues in music information retrieval: ISMIR 2000–2002," *Journal of New Music Research*, vol. 32, no. 2 pp. 121–131, 2003.

[37] J. S. Downie, "The music information retrieval evaluation exchange (2005–2007): A window into music information retrieval research," *Acoustical Science and Technology*, vol. 29, no. 4 pp. 247–255, 2008.

[38] P. Smaragdis, B. Raj, and M. Shashanka, "Supervised and semisupervised separation of sounds from single-channel mixtures," *in Proc. International Conference on Independent Component Analysis and Signal Separation*, 2007, pp. 414–421.

[39] D. Kitamura, H. Saruwatari, K. Yagi, K. Shikano, Y. Takahashi, and K. Kondo, "Music signal separation based on supervised non-negative matrix factorization with orthogonal-ity and maximum-divergence penalties," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E97-A, no. 5, pp.1113–1118, 2014.

[40] Y. Iwase and D. Kitamura, "Supervised audio source separation based on nonnegative matrix factorization with cosine similarity penalty," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E105-A, no. 6, 2022 (in press).

[41] E. M. Grais, M. U. Sen, and H. Erdogan, "Deep neural networks for single channel source separation," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3734–3738, 2014.

[42] P. S. Huang, M. Kim, M. H. Johnson, and P. Smaragdis, "Joint optimization of masks and deep recurrent neural networks for monaural source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2136–2147, 2015.

[43] J. R. Hershey, Z. Chen, J. Le Roux, and S. Watanabe, "Deep clustering: discriminative embeddings for segmentation and separation," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2016, pp. 31–35.

[44] W. Lim and T. Lee, "Harmonic and percussive source separation using a convolu-

tional auto encoder," *in Proc. 25th European Signal Processing Conference*, 2017, pp. 1804–1808.

[45] K. Drossos, P. Magron, S. I. Mimilakis, and T. Virtanen, "Harmonic-percussive source separation with deep neural networks and phase recovery," *in Proc. 16th International Workshop on Acoustic Signal Enhancement*, 2018, pp. 421–425.

[46] N. Q. K. Duong, H. Tachibana, E. Vincent, N. Ono, R. Gribonval, and S. Sagayama, "Multichannel harmonic and percussive component separation by joint modeling of spatial and spectral continuity," *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 205–208.

[47] N. Q. K. Duong, E. Vincent and R. Gribonval, "Underdetermined reverberant audio source separation using local observed covariance and auditory-motivated timefrequency representation," *in Proc. International Conference on Latent Variable Analysis and Signal Separation*, 2010, pp. 73–80.

[48] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Underdetermined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, no. 7, pp. 1830–1840, 2010.

[49] T. K. Moon, "The expectation-maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[50] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1–4, pp. 1–24, 2001.

[51] K. Matsuoka, "Minimal distortion principle for blind source separation," *in Proc. Proceedings of the 41st Society of Instrument and Control Engineers Annual Conference.*, 2002, vol. 4, pp. 2138–21437.

[52] A. Liutkus, F.-R. Stöter, Z. Rafii, D. Kitamura, B. Rivet, N. Ito, N. Ono, and J. Fontecave, "The 2016 signal separation evaluation campaign," *in Proc. Latent Variable Analysis Signal Separation*, 2017, pp. 323–332.

[53] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," *in Proc. International Conference on Language Resources and Evaluation*, 2000, pp. 965–968.

[54] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 4, pp. 1462–1469, 2006.

発表文献一覧

査読付き国際会議

1. **Soichiro Oyabu**, Daichi Kitamura, and Kohei Yatabe, “Linear multichannel blind source separation based on time-frequency mask obtained by harmonic/percussive sound separation,” *in Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2021)*, pp. 201–205, Tronto, Canada, June 2021.

国内学会

1. **大藪宗一郎**, 北村大地, 矢田部浩平, “メディアン型 HPSS を用いた時間周波数マスクに基づくブラインド音源分離, 日本音響学会 2021 年春季研究発表会講演論文集, 2-1Q-18, pp. 411–414, 東京, 2021 年 3 月.
2. **大藪宗一郎**, 北村大地, 矢田部浩平, “調波打撃音分離の排他的マスキングに基づくブラインド音源分離, 日本音響学会 2020 年秋季研究発表会講演論文集, 2-R2-11, pp. 283–286, 宮城, 2020 年 9 月.
3. **大藪宗一郎**, 北村大地, 矢田部浩平, “調波打撃音分離の時間周波数マスクを用いた線形ブラインド音源分離, 日本音響学会 2020 年春季研究発表会講演論文集, 3-1-16, pp. 313–316, 埼玉, 2020 年 3 月.

付録 A

OHPSS モデルの各楽曲における SDR 改善量の収束挙動

OHPSS モデルの各楽曲における SDR 改善量の収束挙動を Figs. A.1–A.19 に示す．Figs. A.1–A.19 は，4.5 節で例に挙げた楽曲（song no. 1）以外の楽曲（song nos. 2–20）に対応する実験結果である．

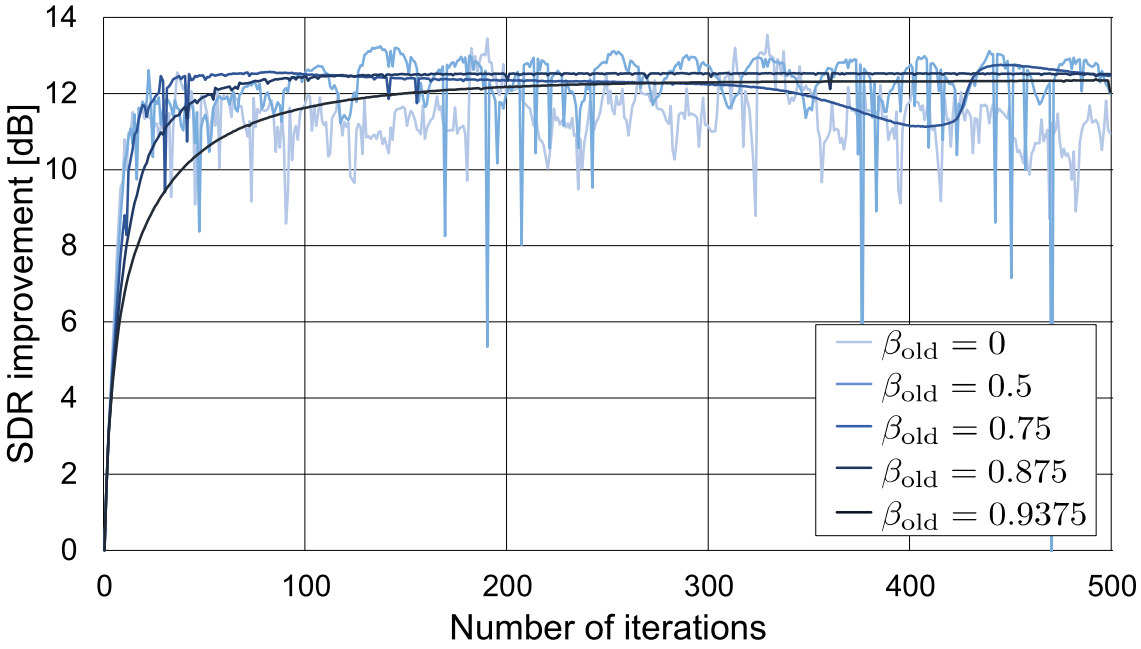

Fig. A.1. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 2).

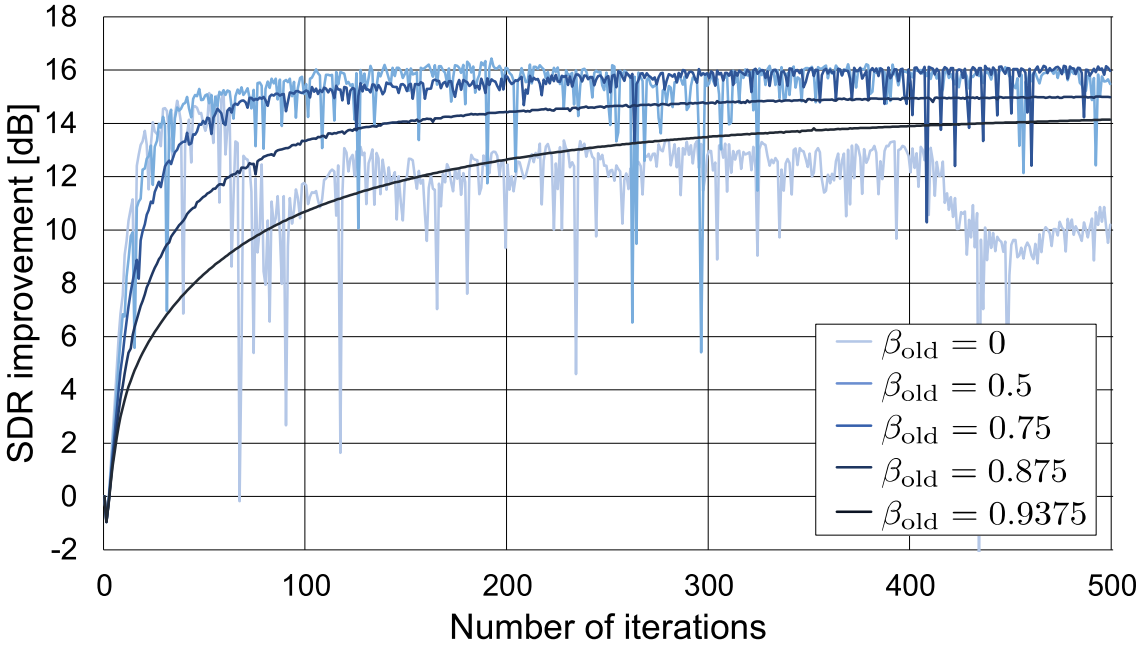

Fig. A.2. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 3).

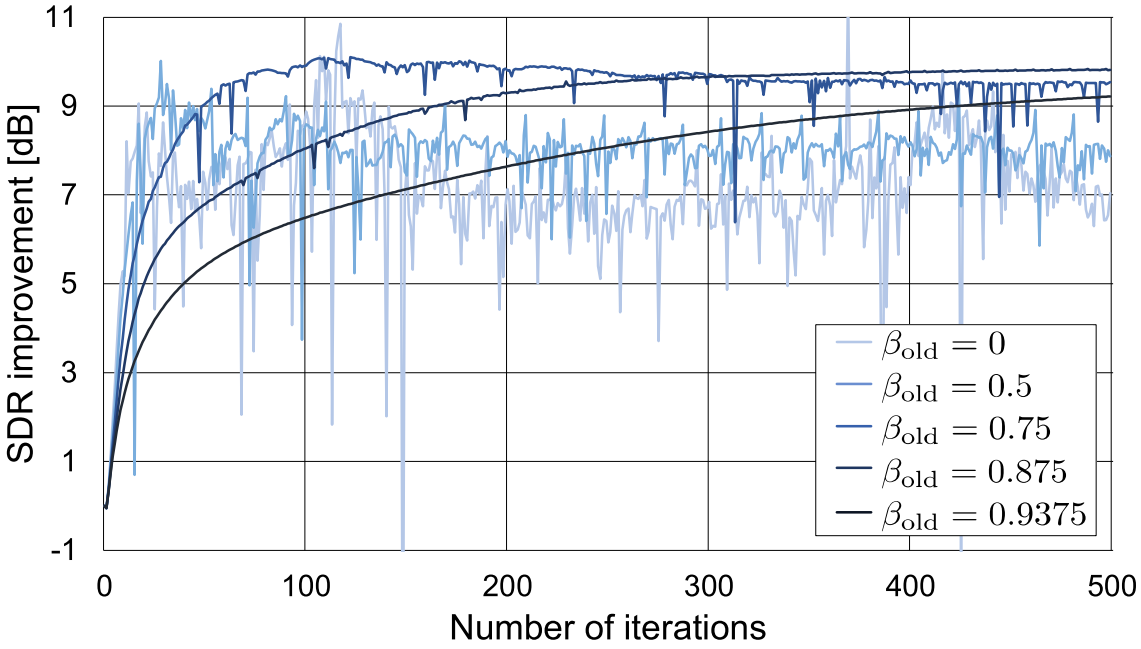

Fig. A.3. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 4).

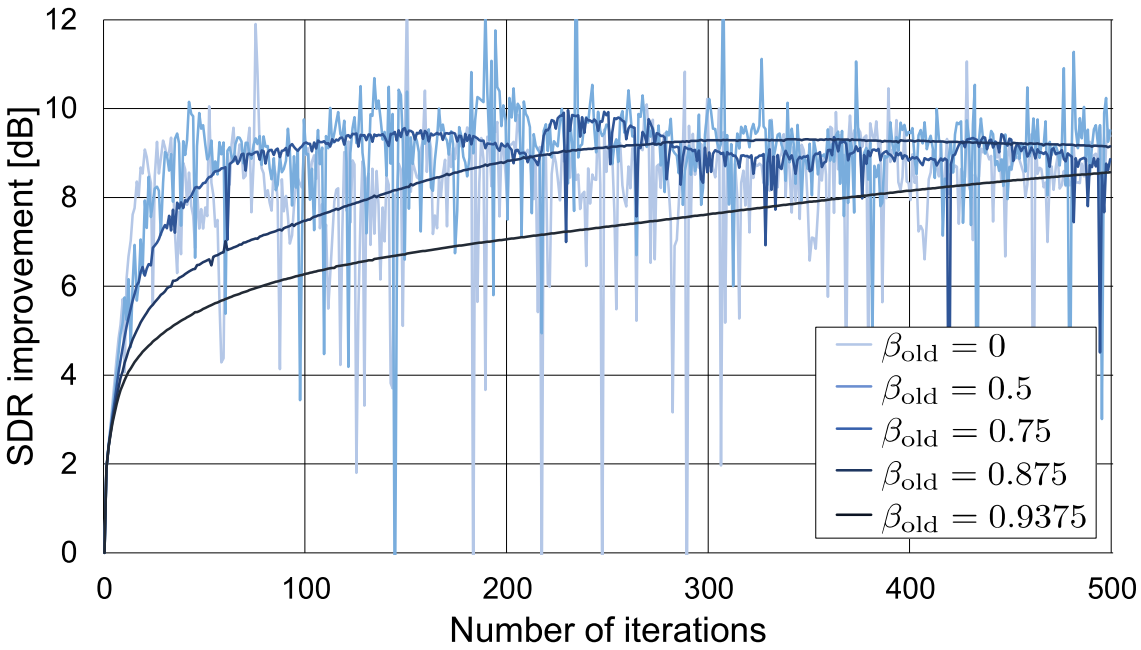

Fig. A.4. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 5).

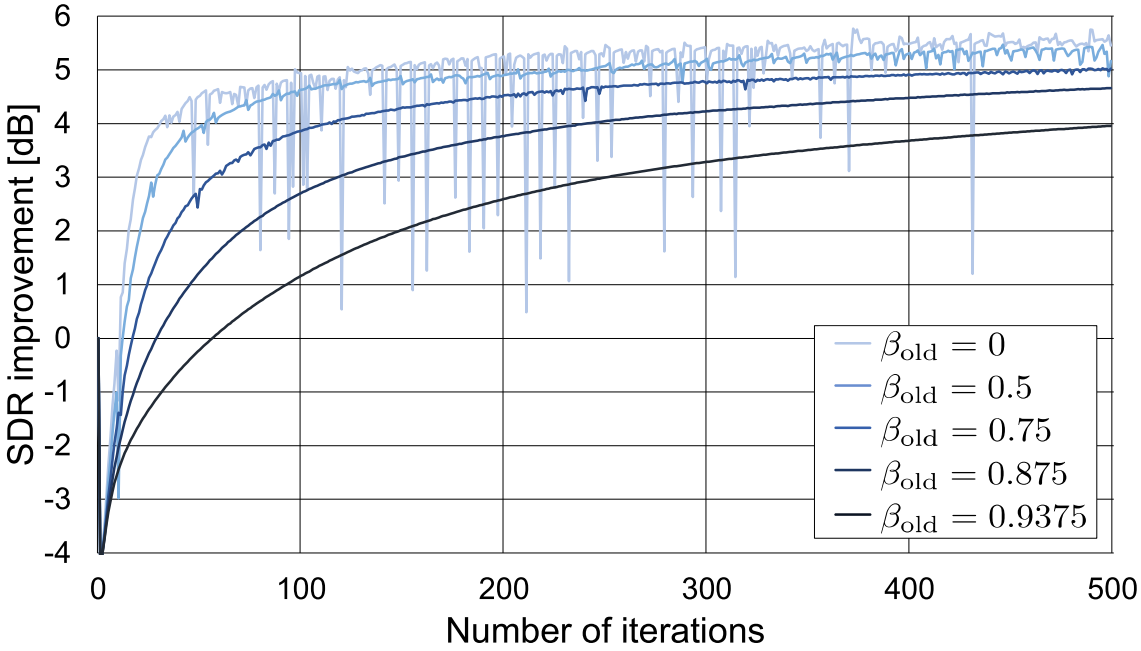

Fig. A.5. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 6).

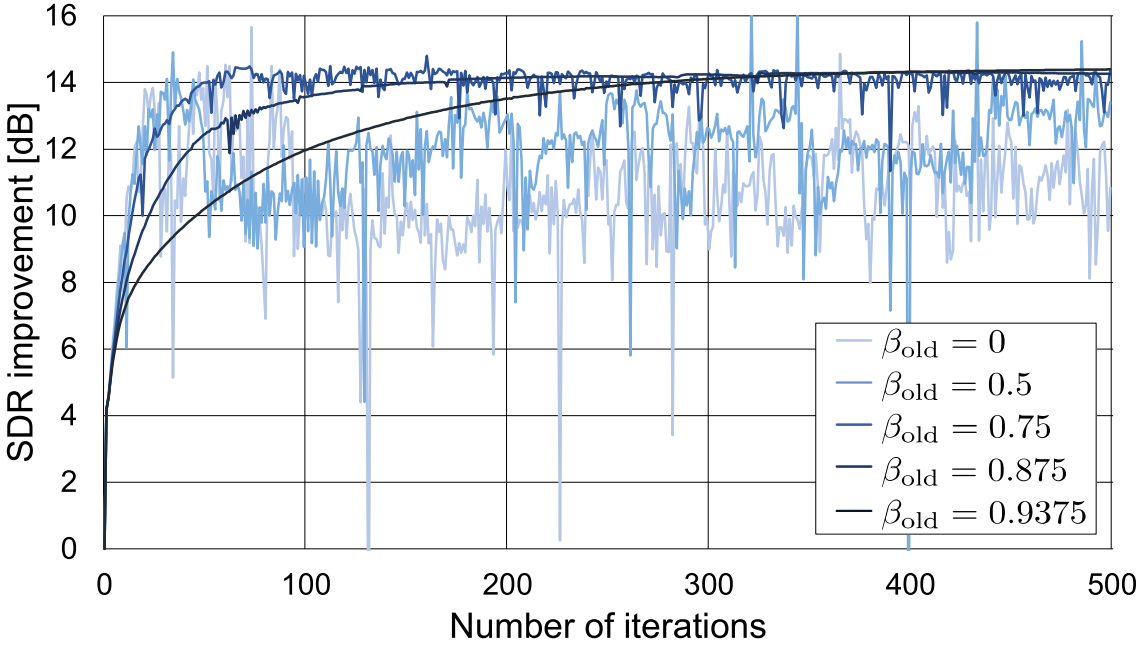

Fig. A.6. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 7).

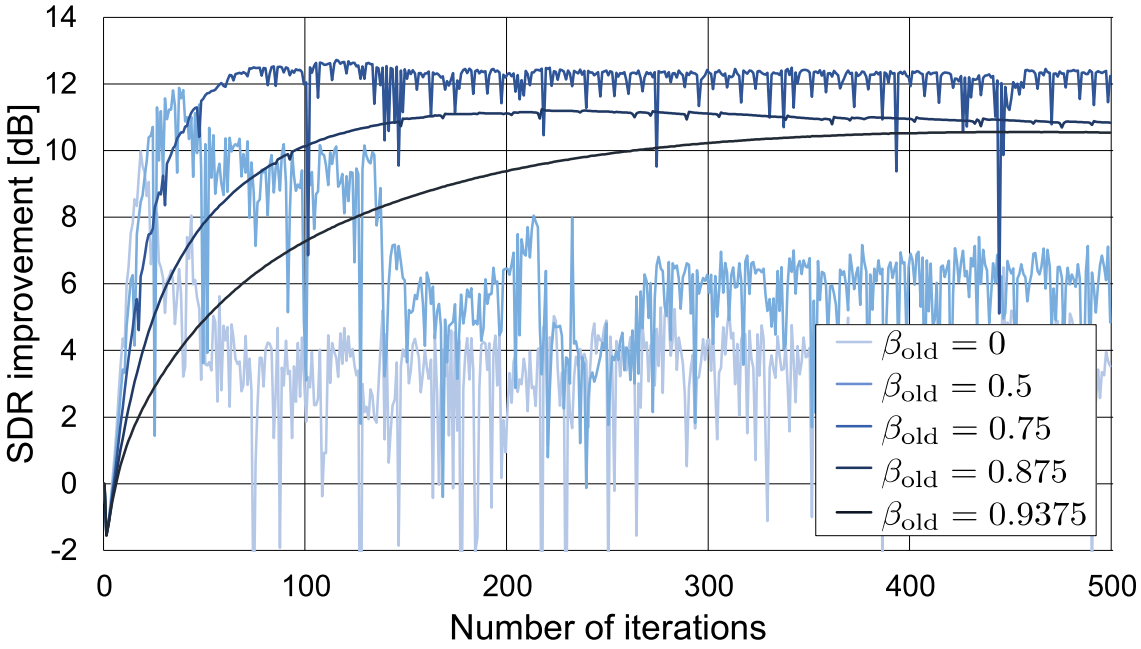

Fig. A.7. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 8).

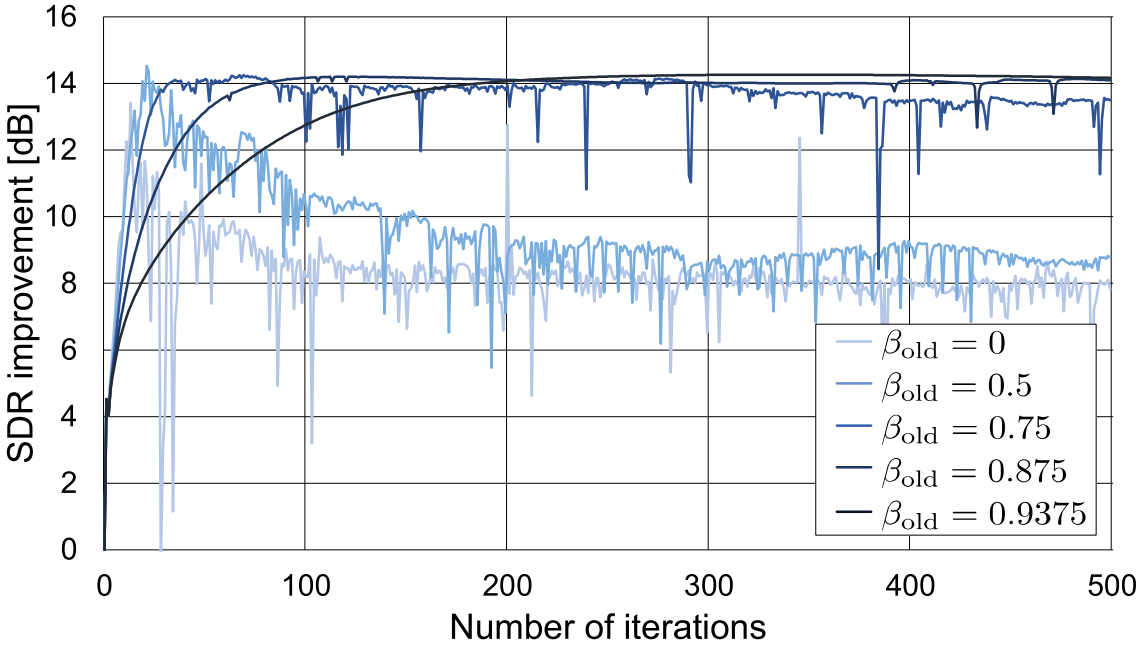

Fig. A.8. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 9).

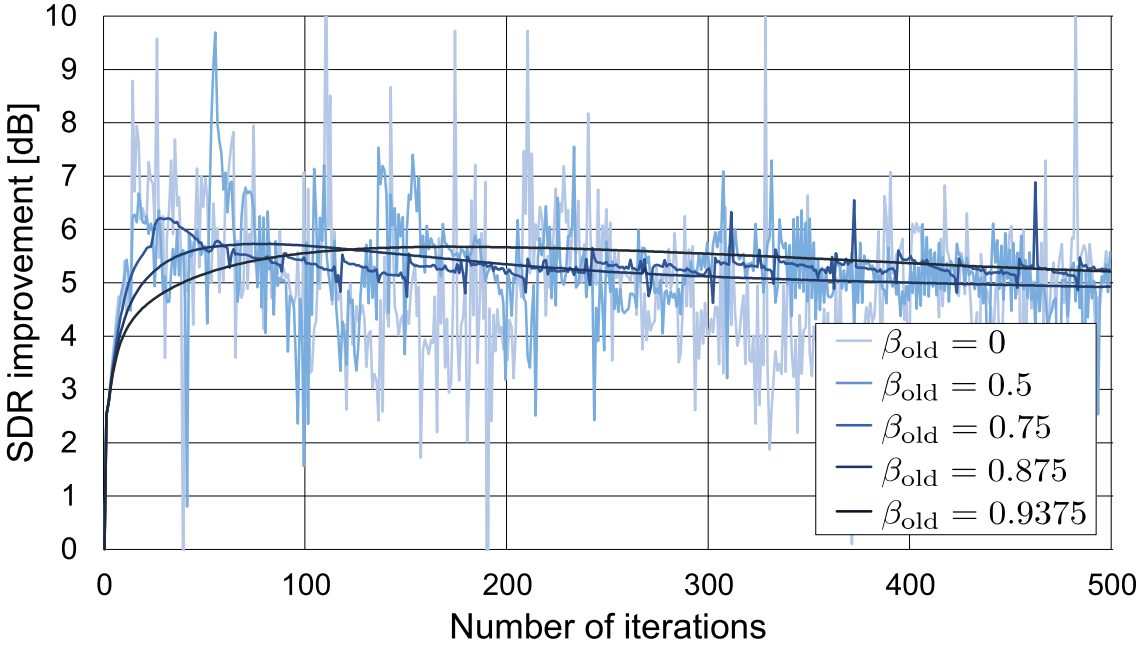

Fig. A.9. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 10).

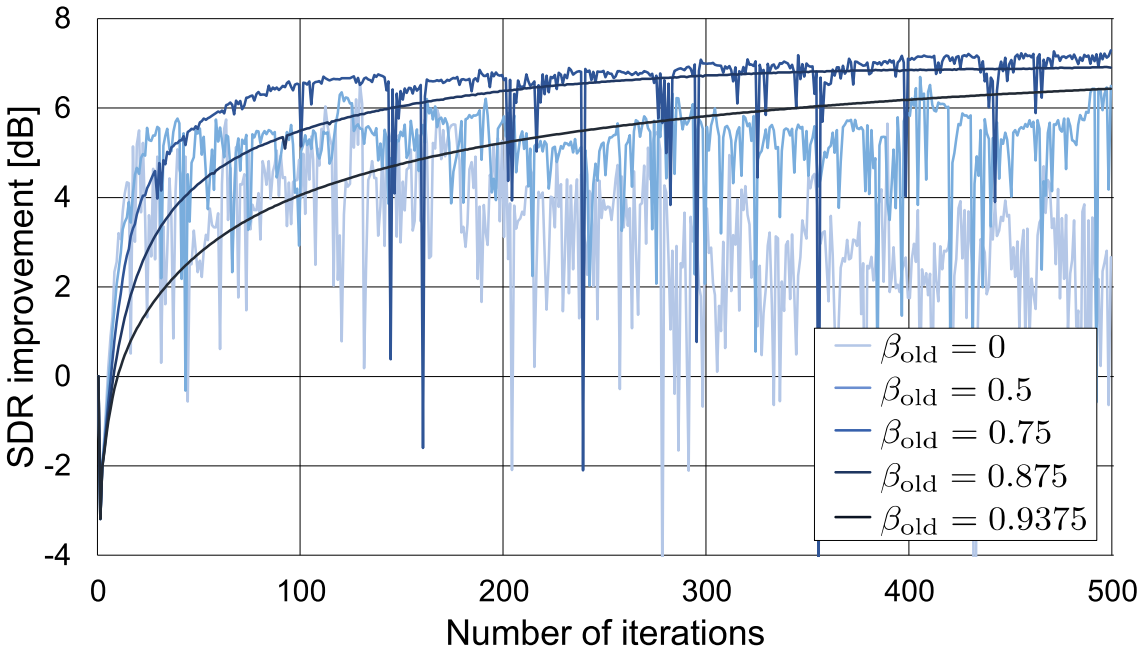

Fig. A.10. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 11).

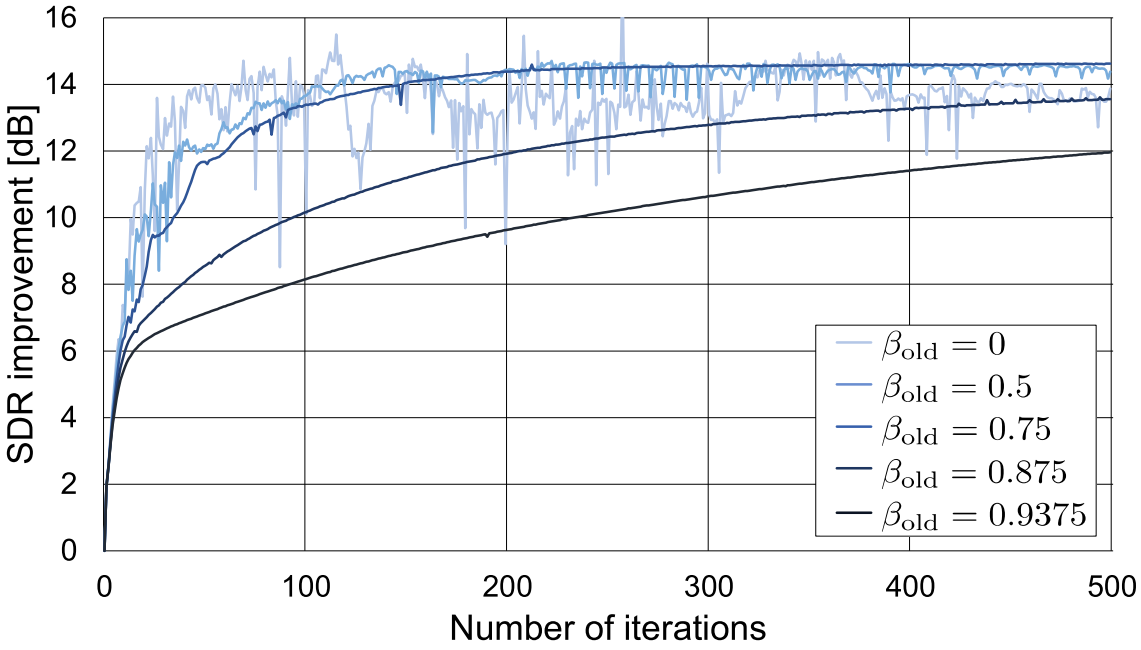

Fig. A.11. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 12).

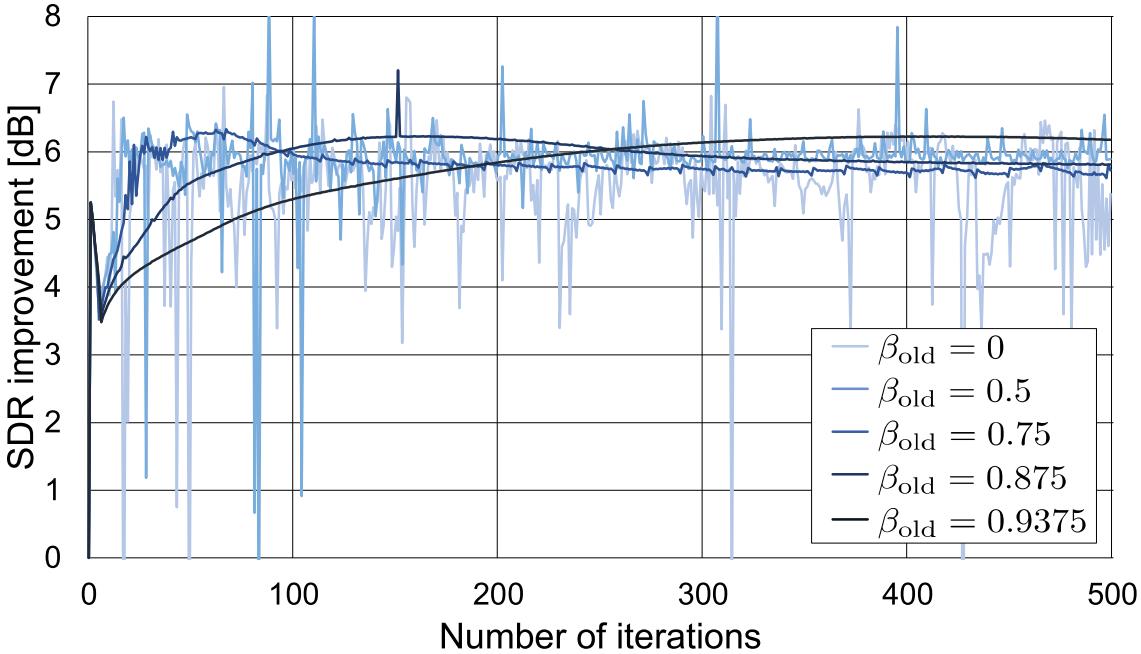

Fig. A.12. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 13).

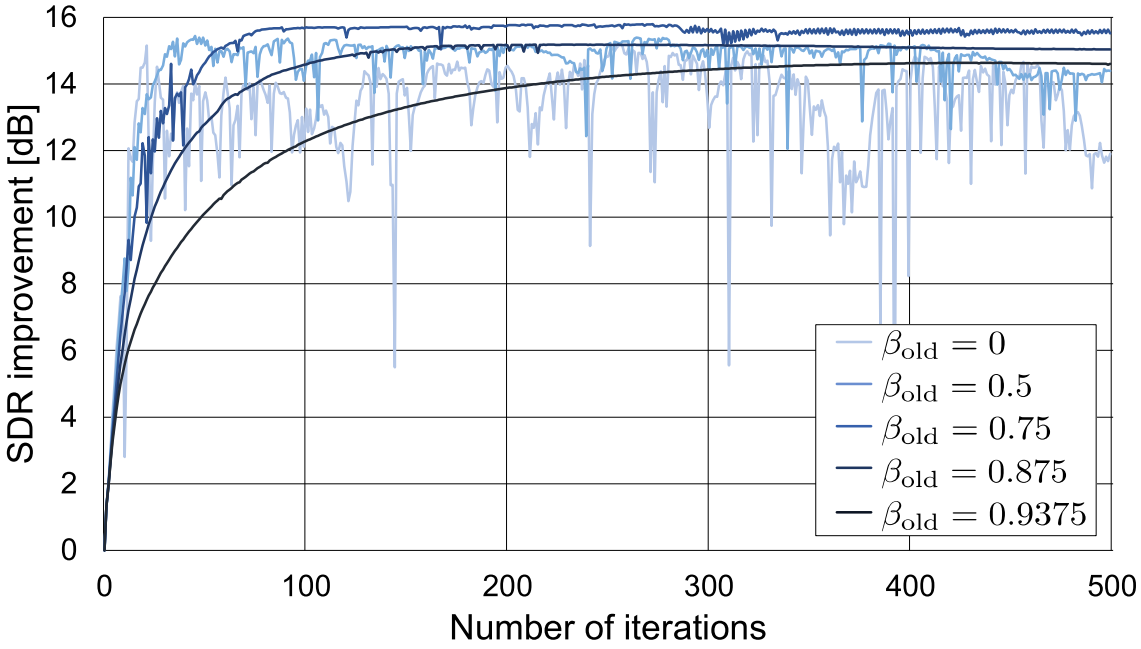

Fig. A.13. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 14).

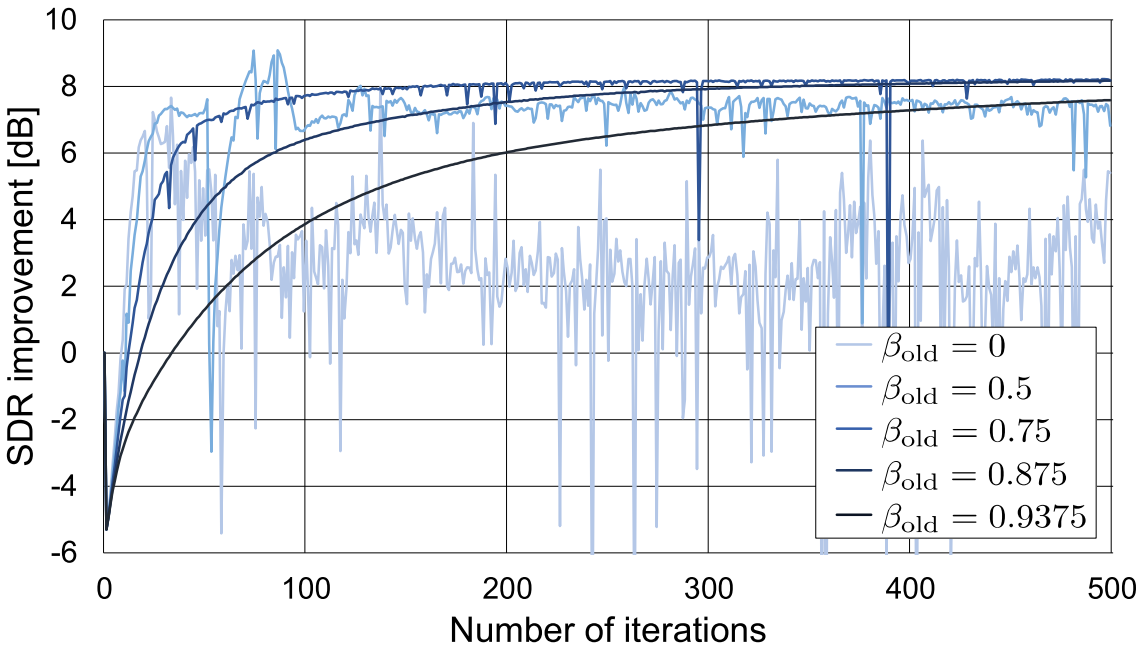

Fig. A.14. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 15).

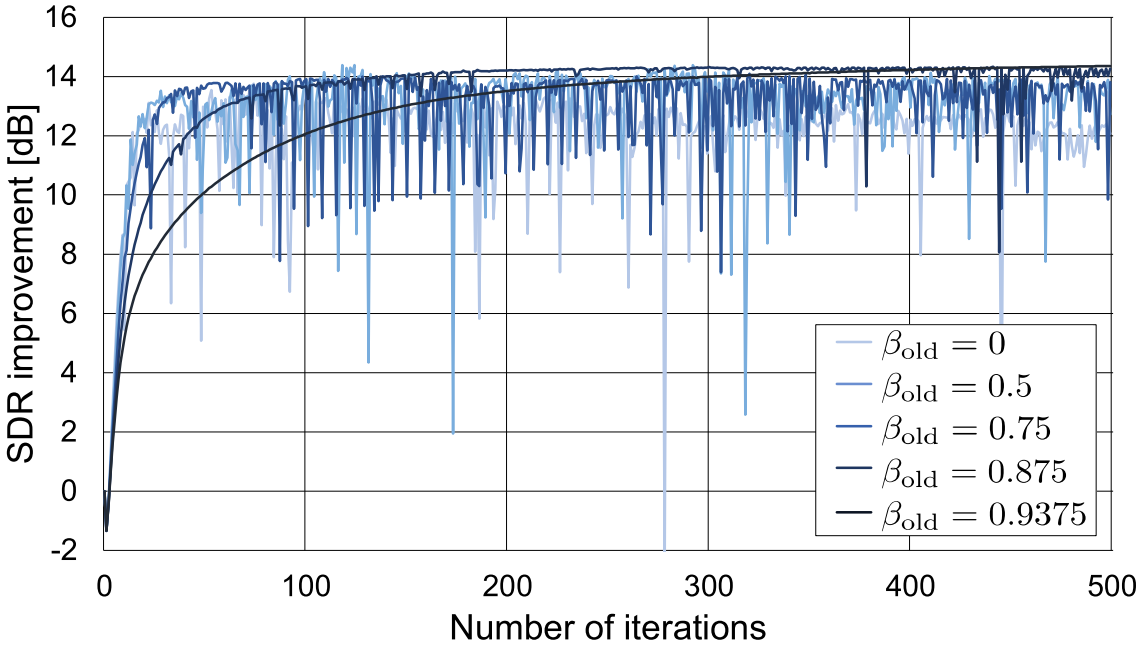

Fig. A.15. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 16).

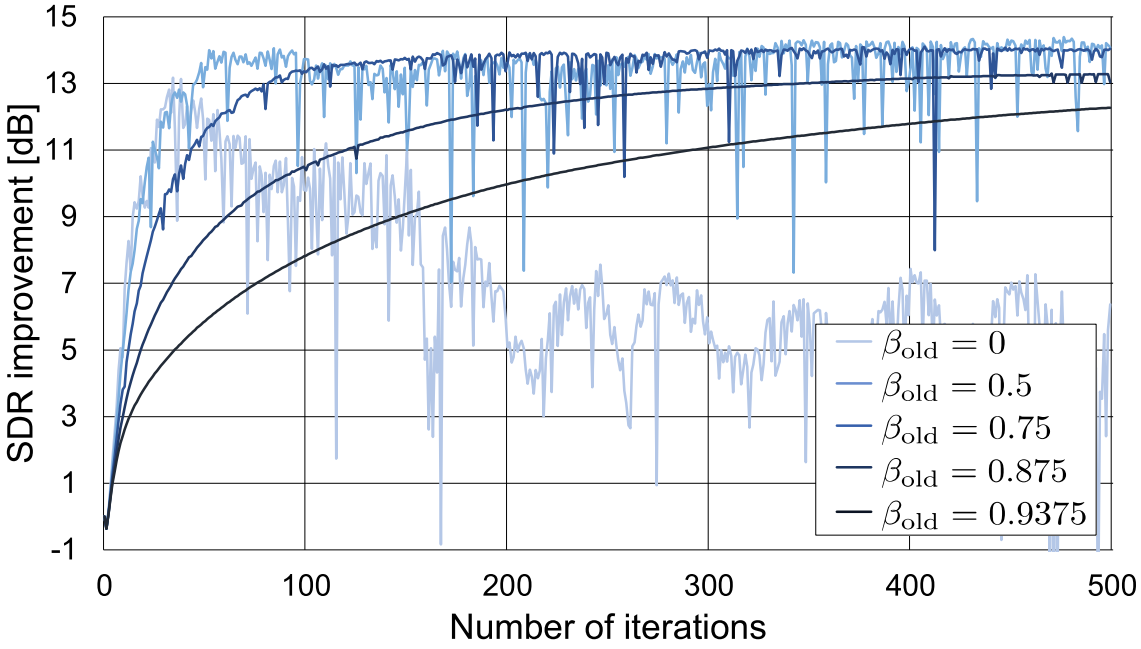

Fig. A.16. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 17).

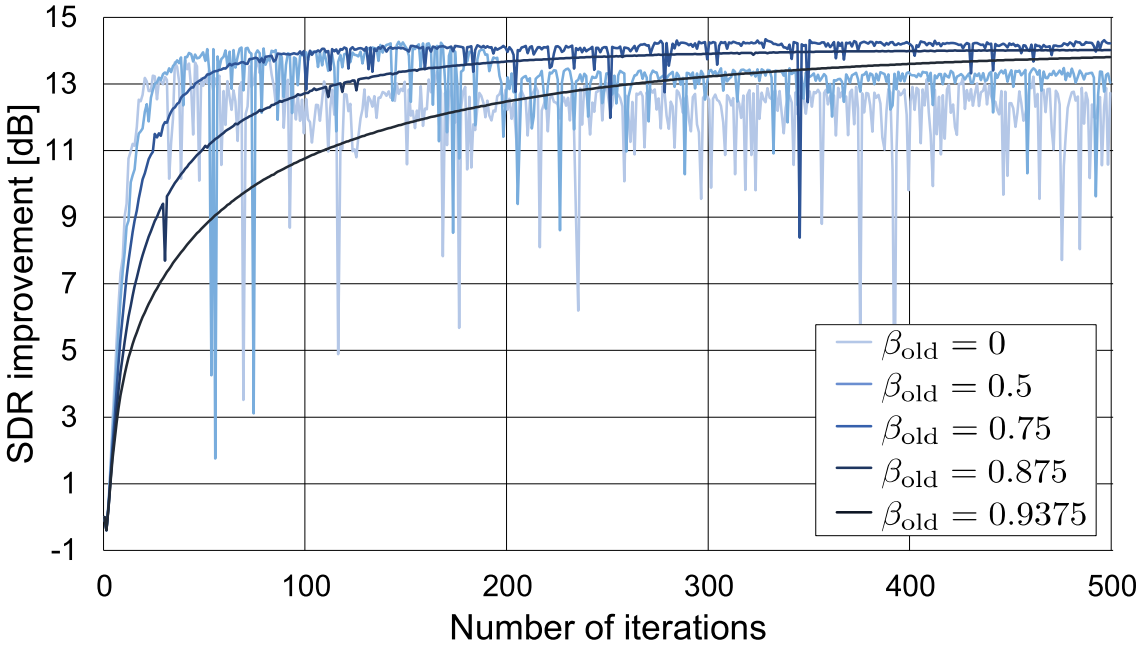

Fig. A.17. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 18).

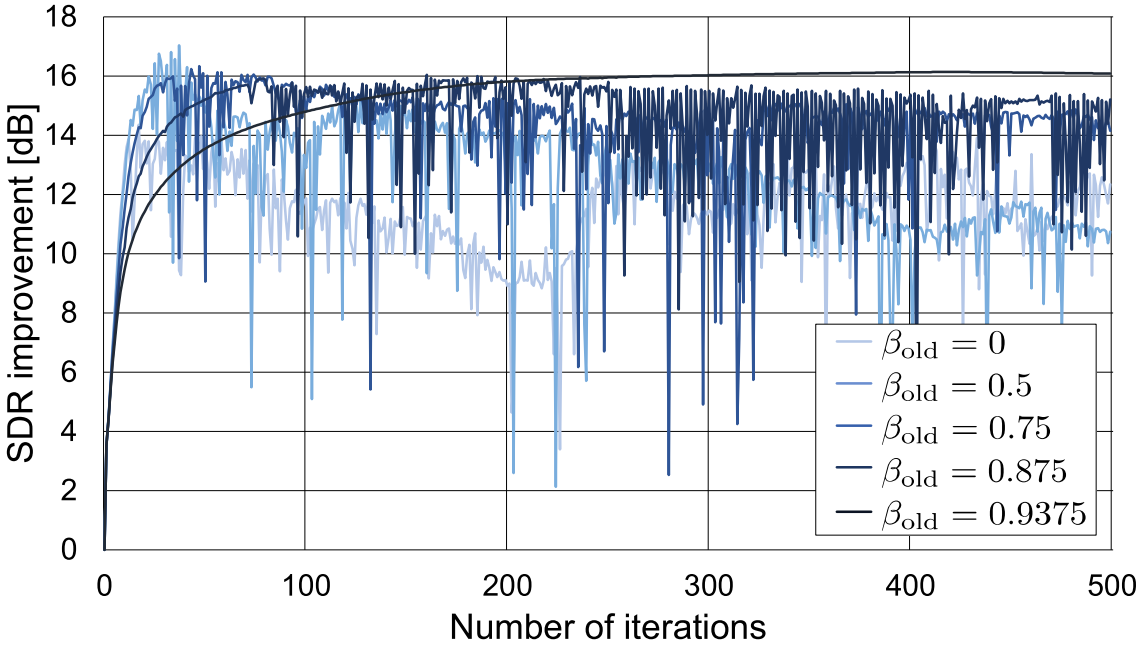

Fig. A.18. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 19).

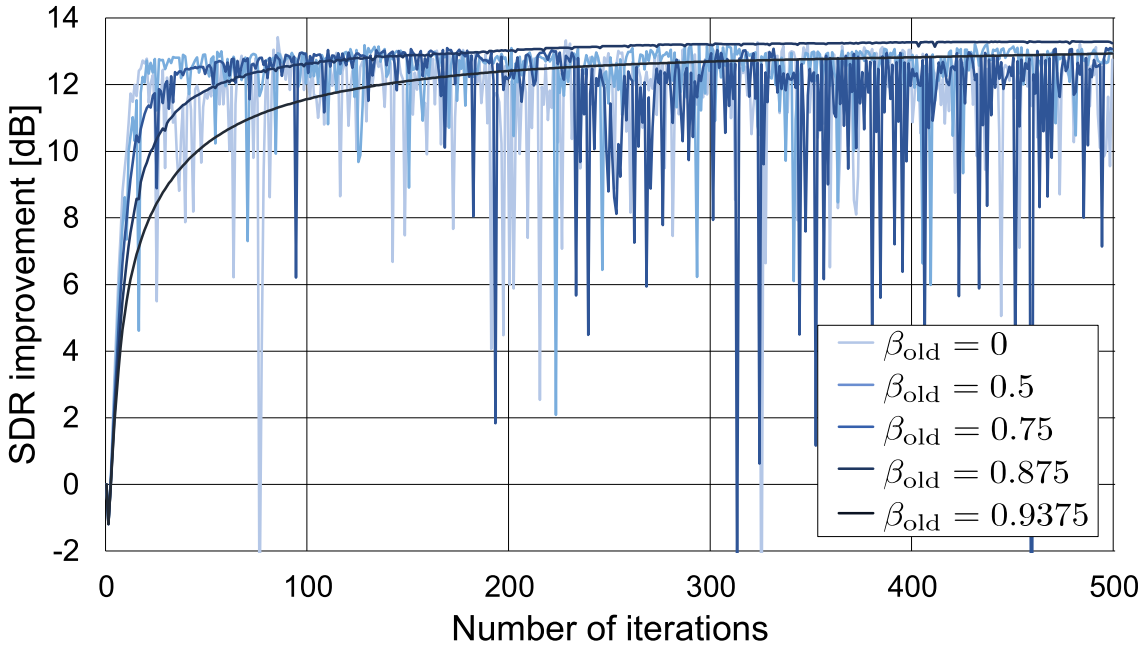

Fig. A.19. Example of convergence behaviors of the OHPSS-based proposed method with various β_{old} and β (song no. 20).

付録 B

MHPSS モデルの各楽曲における SDR 改善量の収束挙動

MHPSS モデルの各楽曲における SDR 改善量の収束挙動を Figs. B.1–B.19 に示す．Figs. B.1–B.19 は，4.5 節で例に挙げた楽曲（song no. 1）以外の楽曲（song nos. 2–20）に対応する実験結果である．

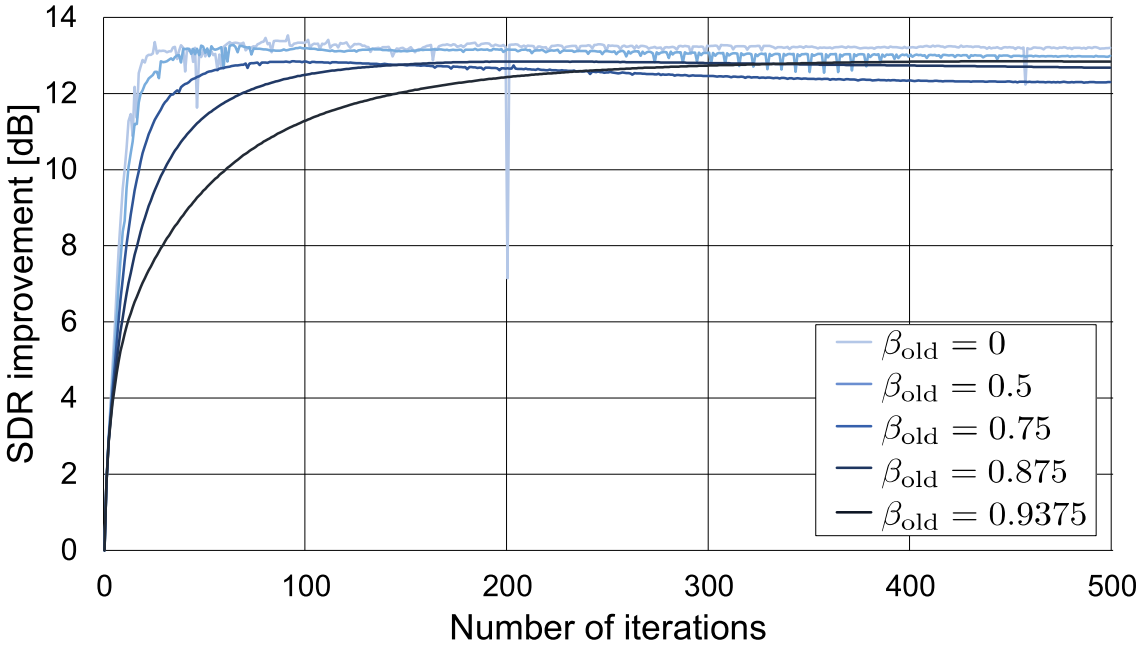

Fig. B.1. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 2).

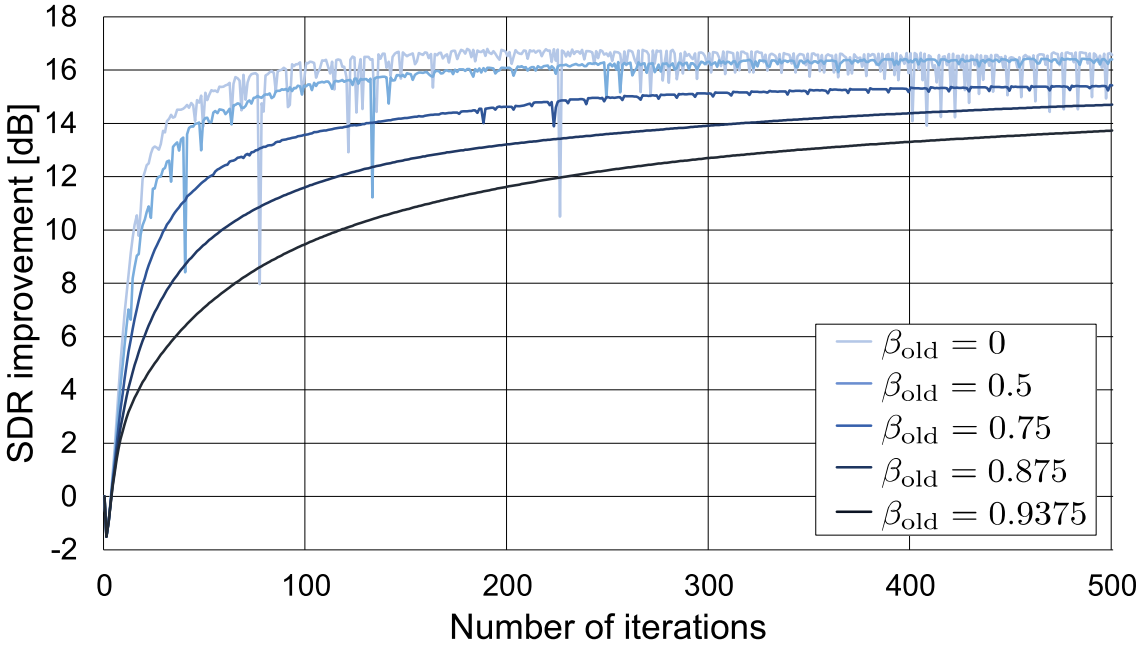

Fig. B.2. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 3).

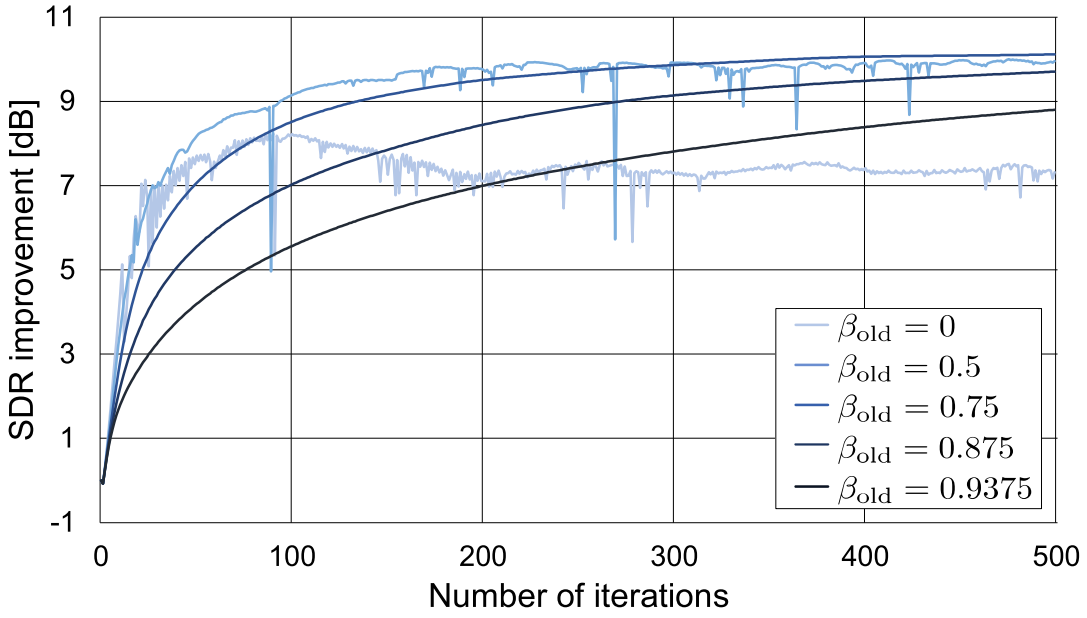

Fig. B.3. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 4).

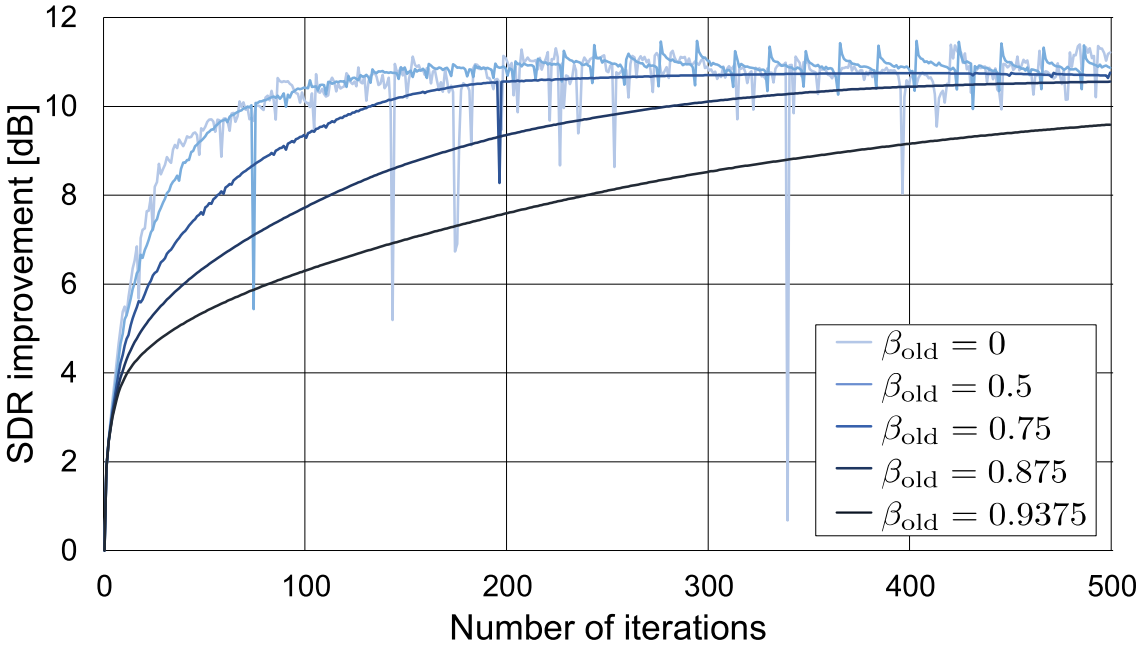

Fig. B.4. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 5).

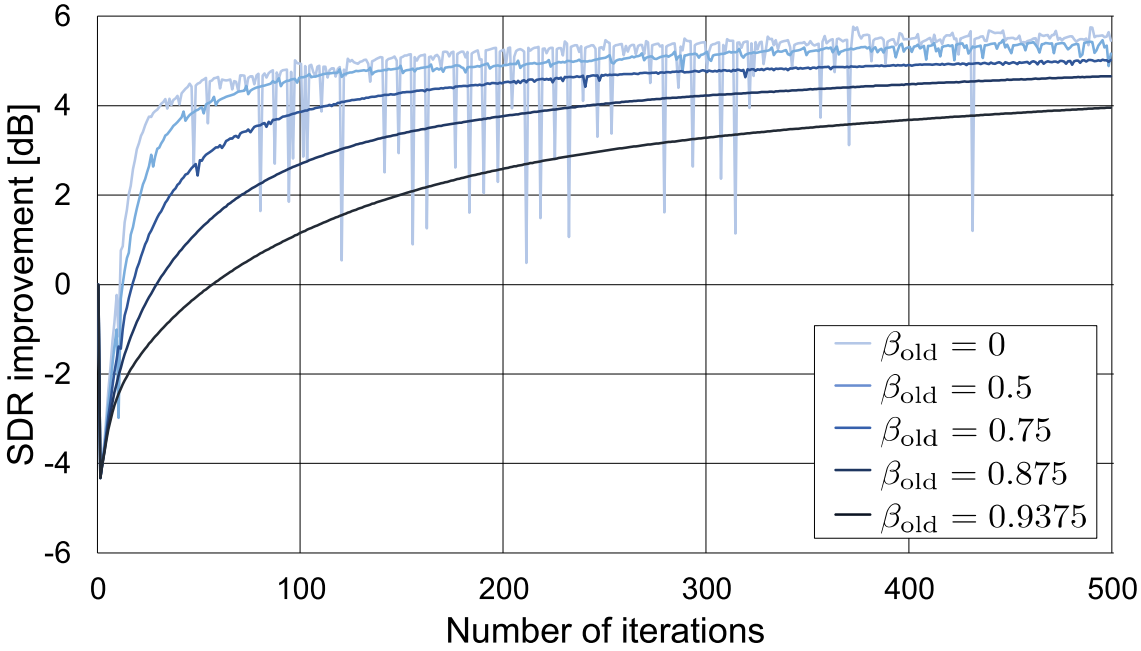

Fig. B.5. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 6).

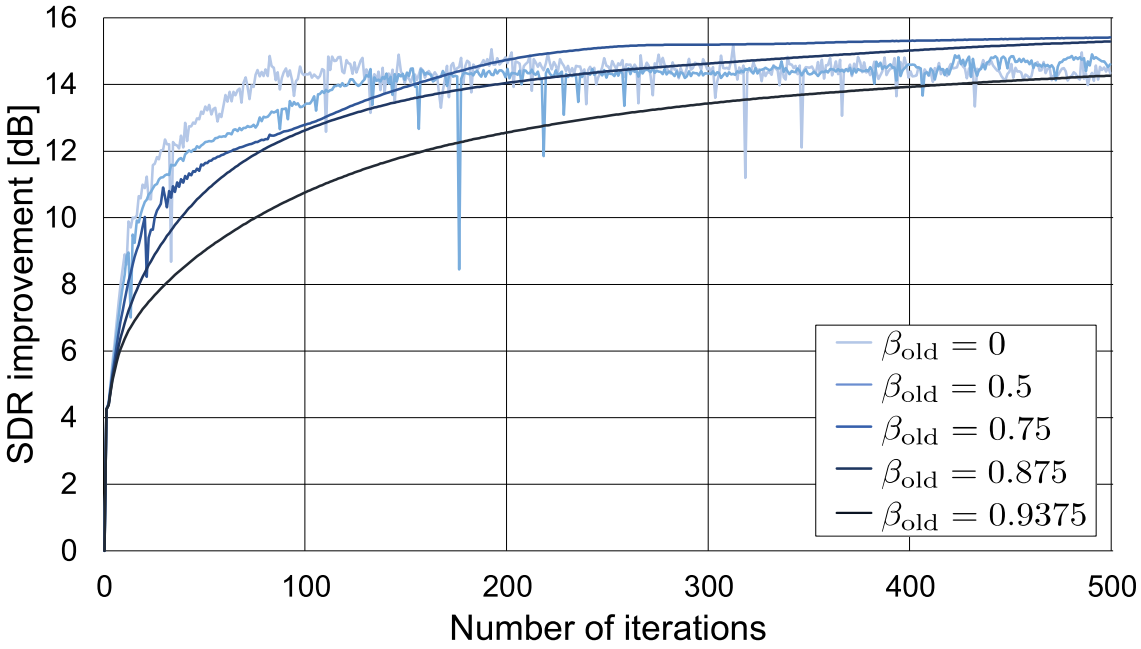

Fig. B.6. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 7).

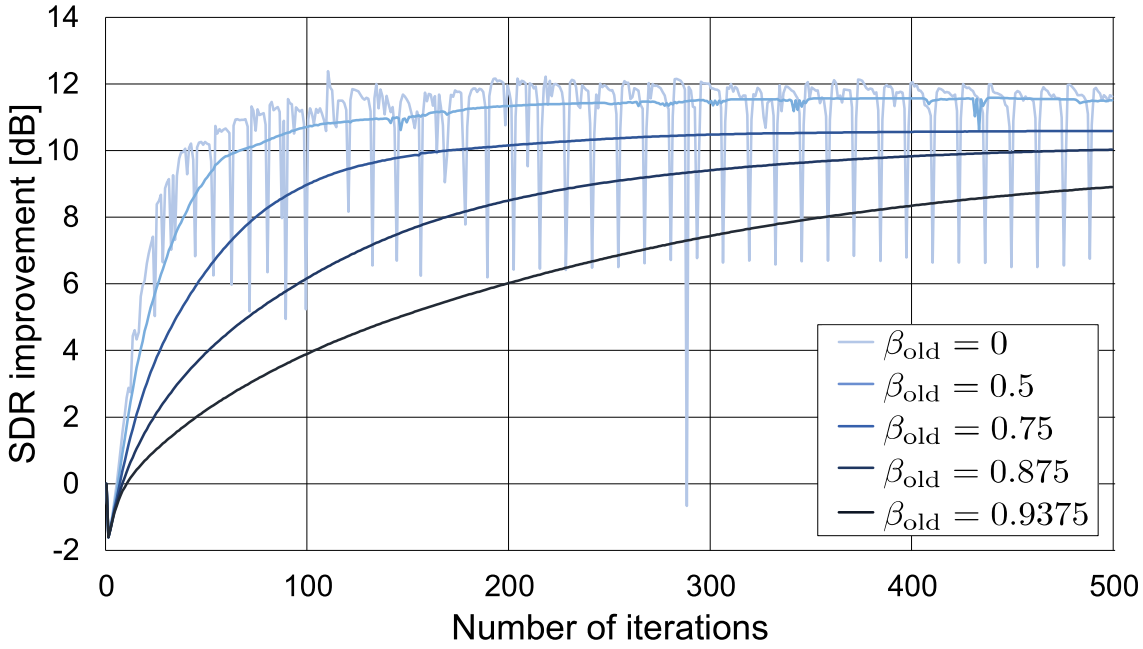

Fig. B.7. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 8).

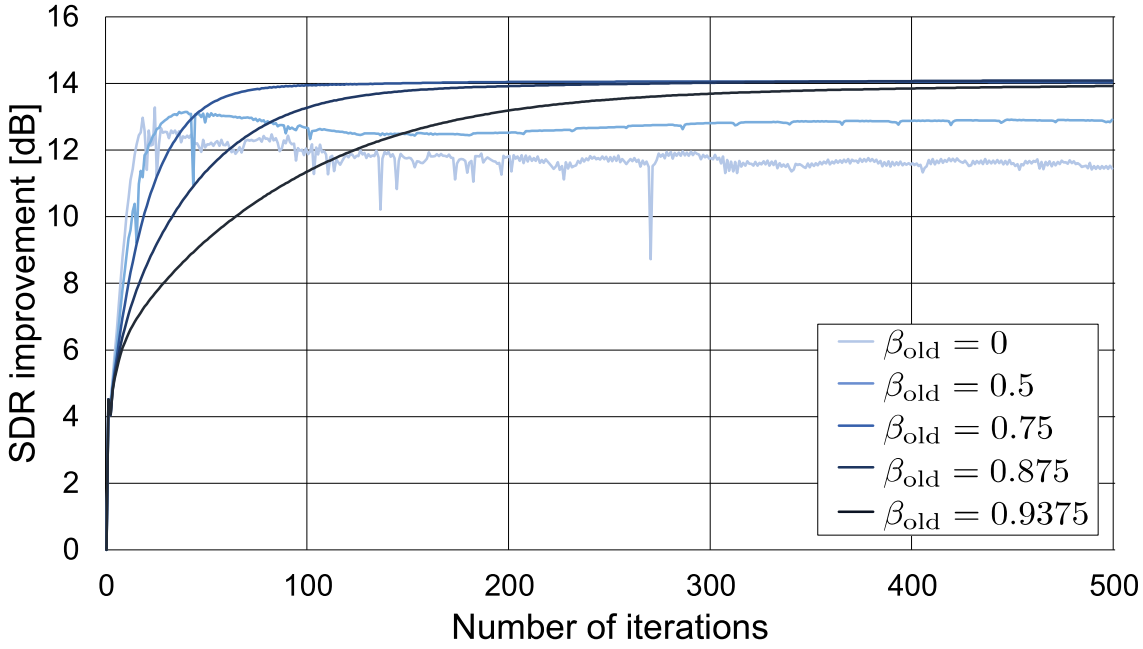

Fig. B.8. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 9).

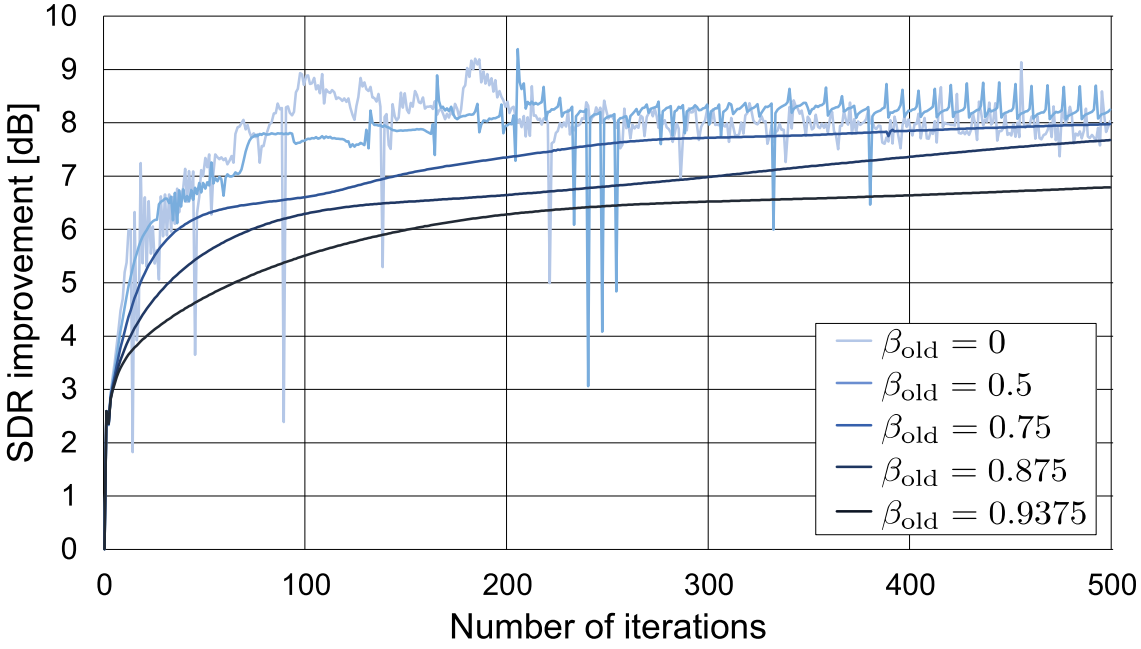

Fig. B.9. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 10).

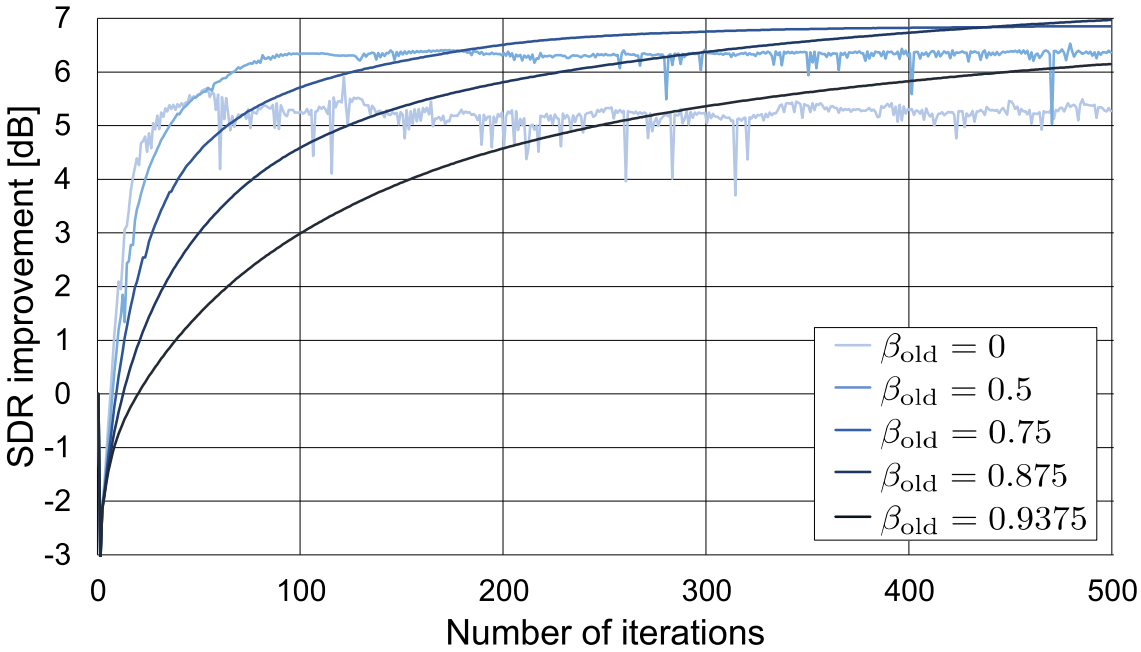

Fig. B.10. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 11).

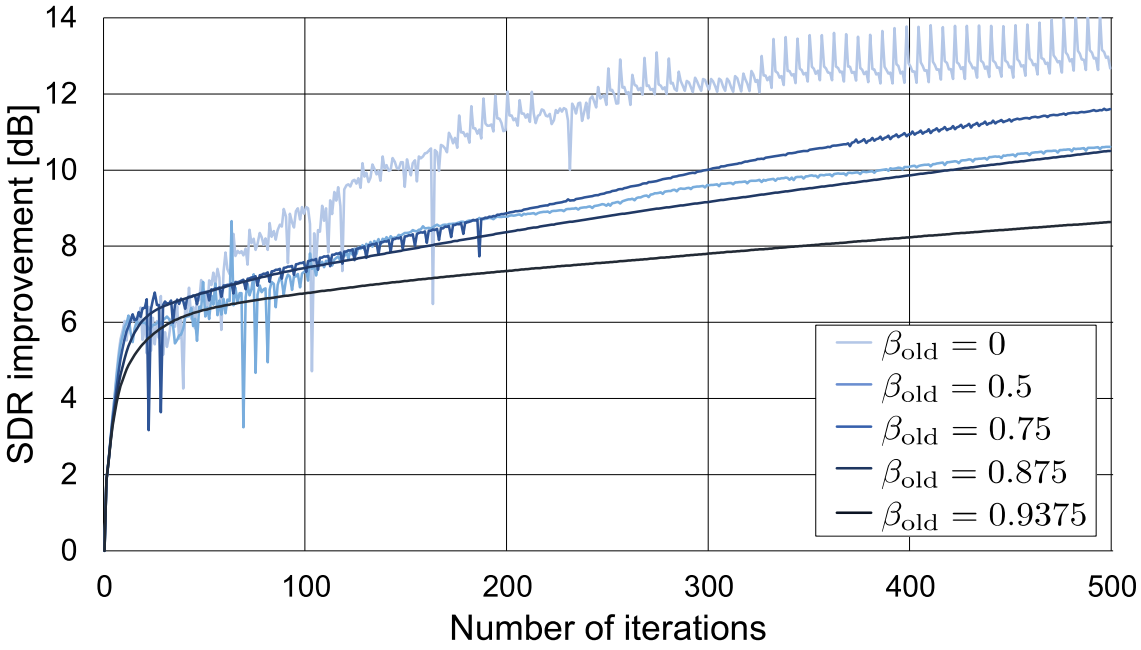

Fig. B.11. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 12).

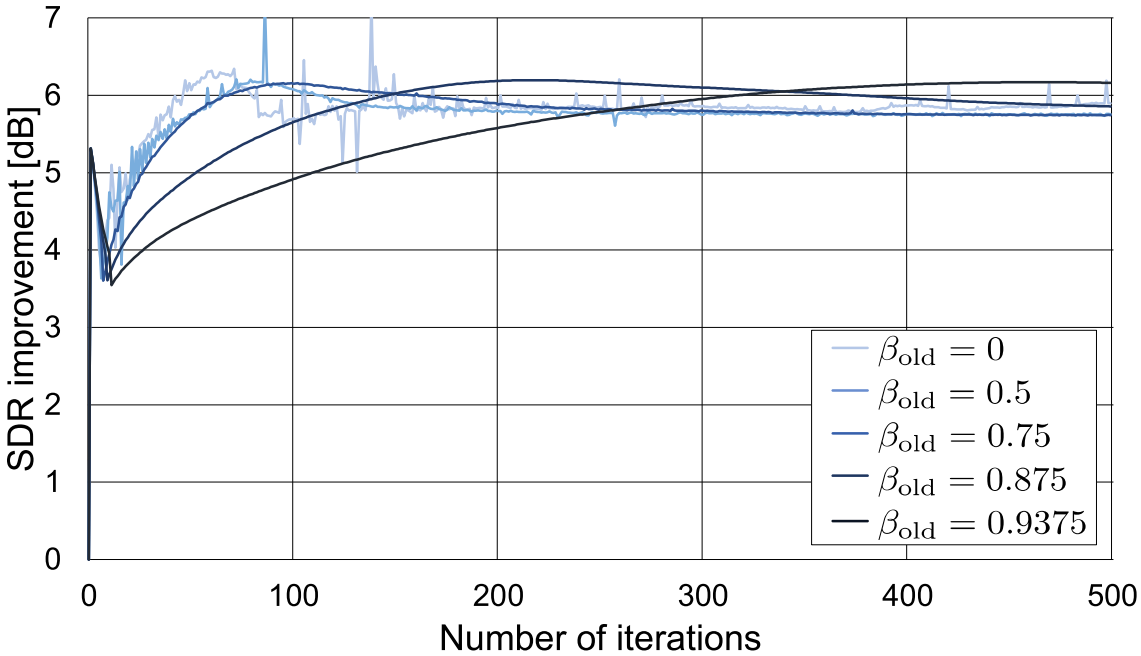

Fig. B.12. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 13).

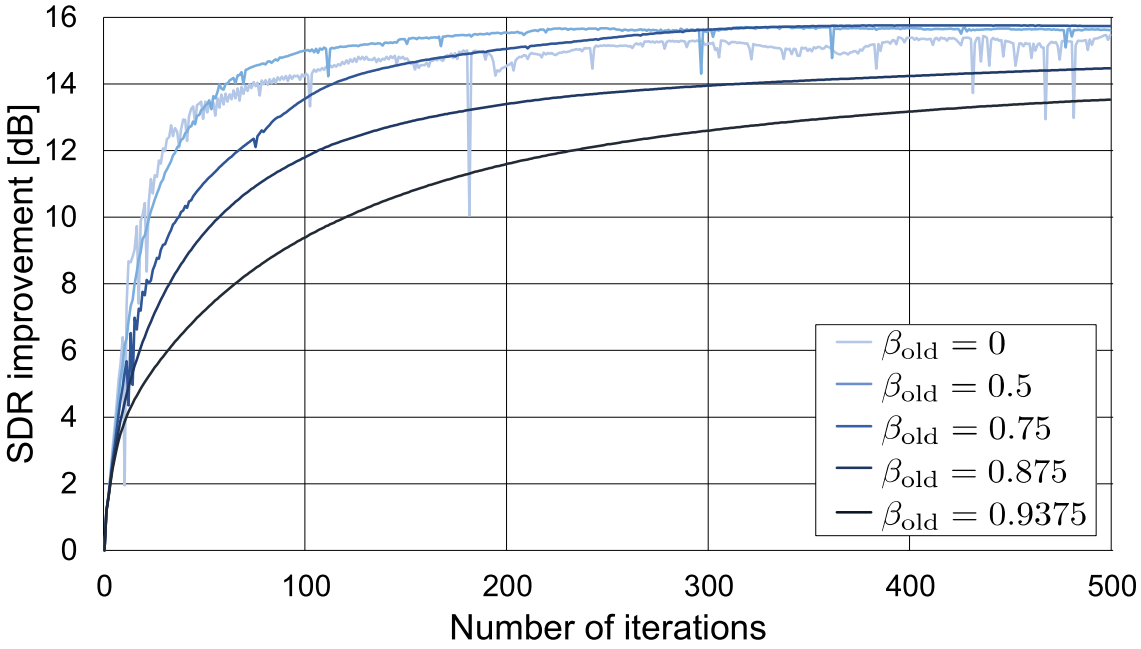

Fig. B.13. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 14).

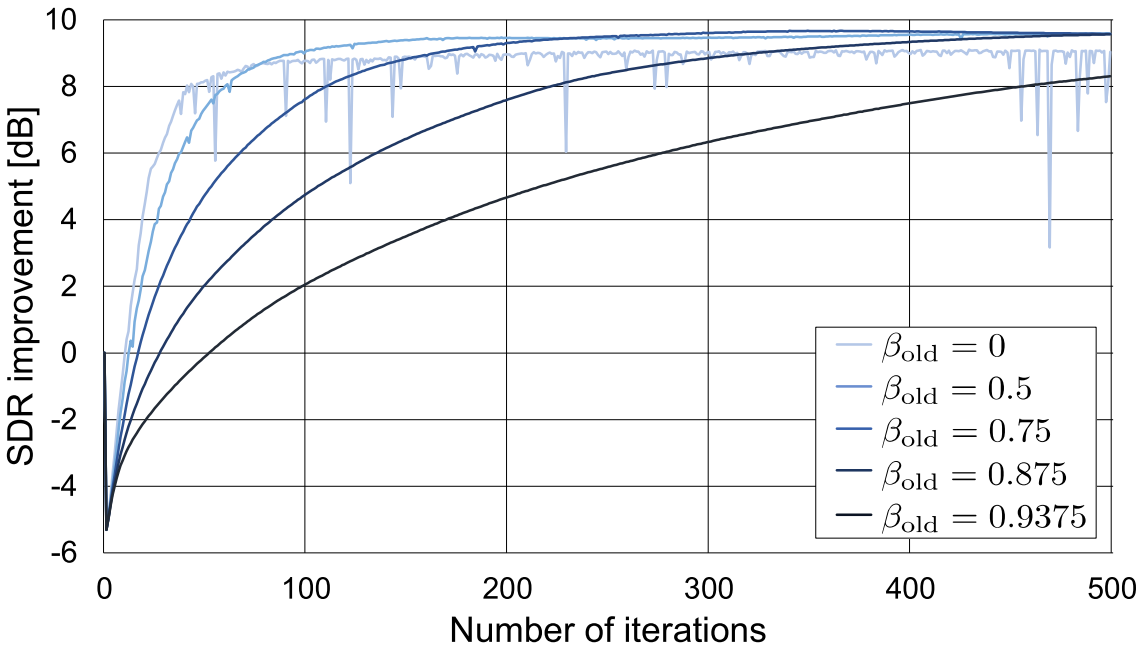

Fig. B.14. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 15).

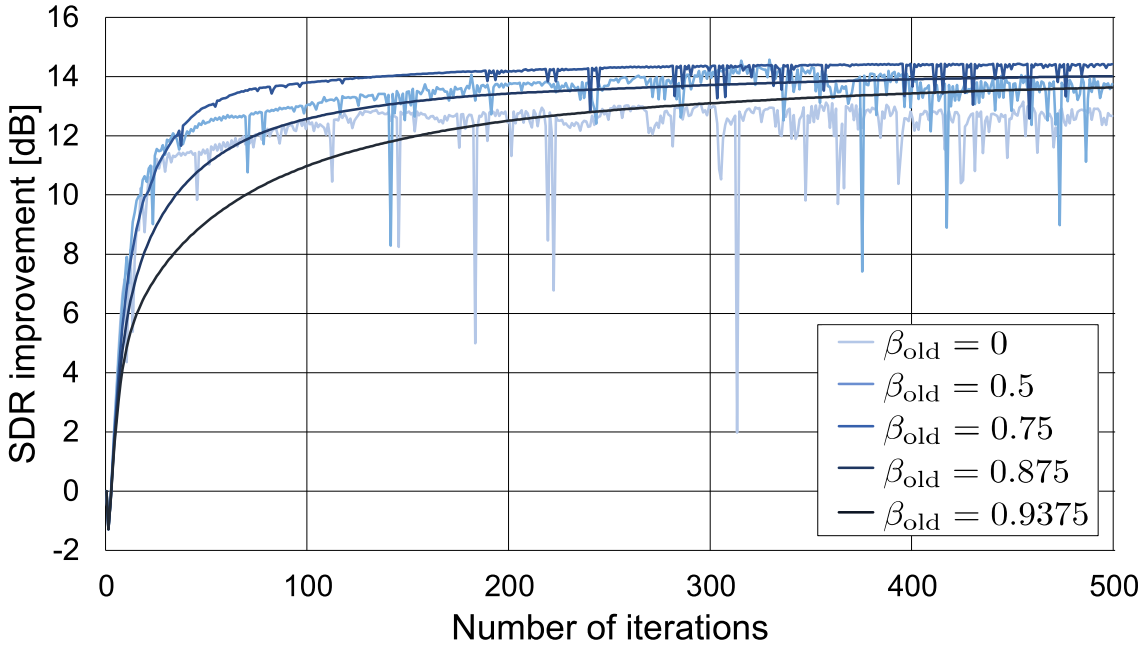

Fig. B.15. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 16).

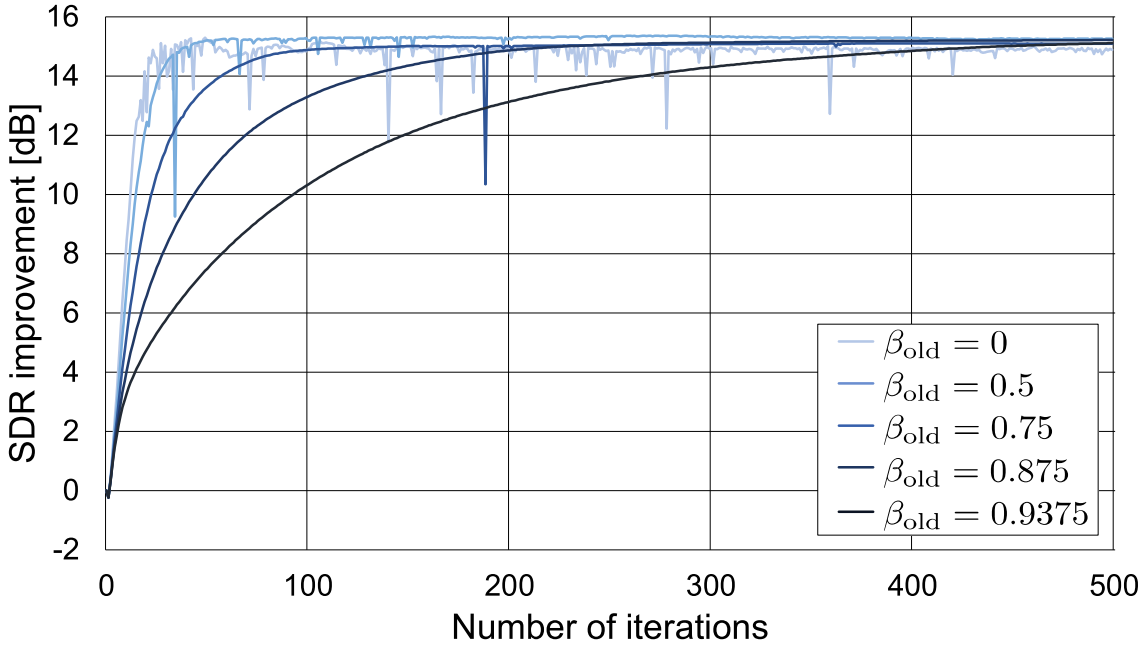

Fig. B.16. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 17).

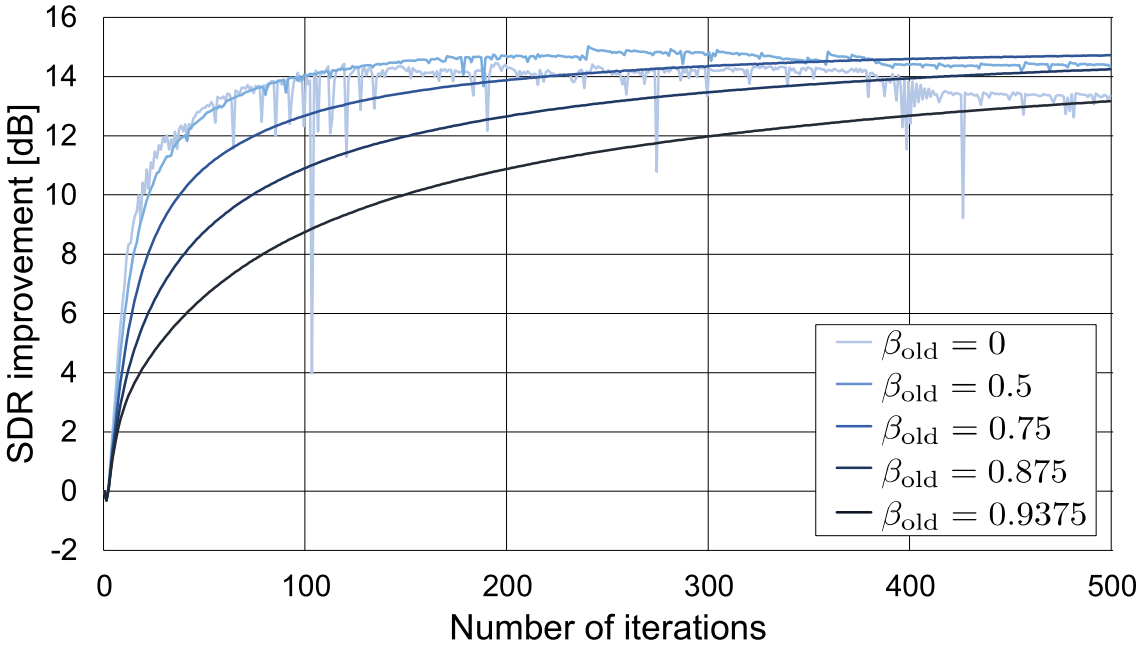

Fig. B.17. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 18).

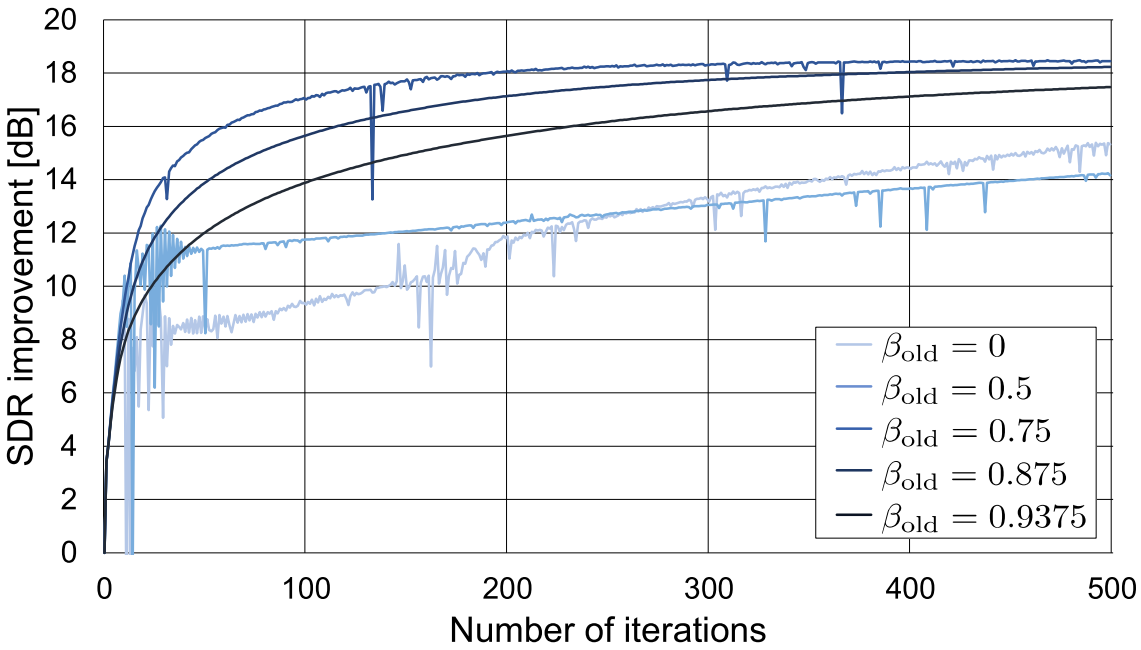

Fig. B.18. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 19).

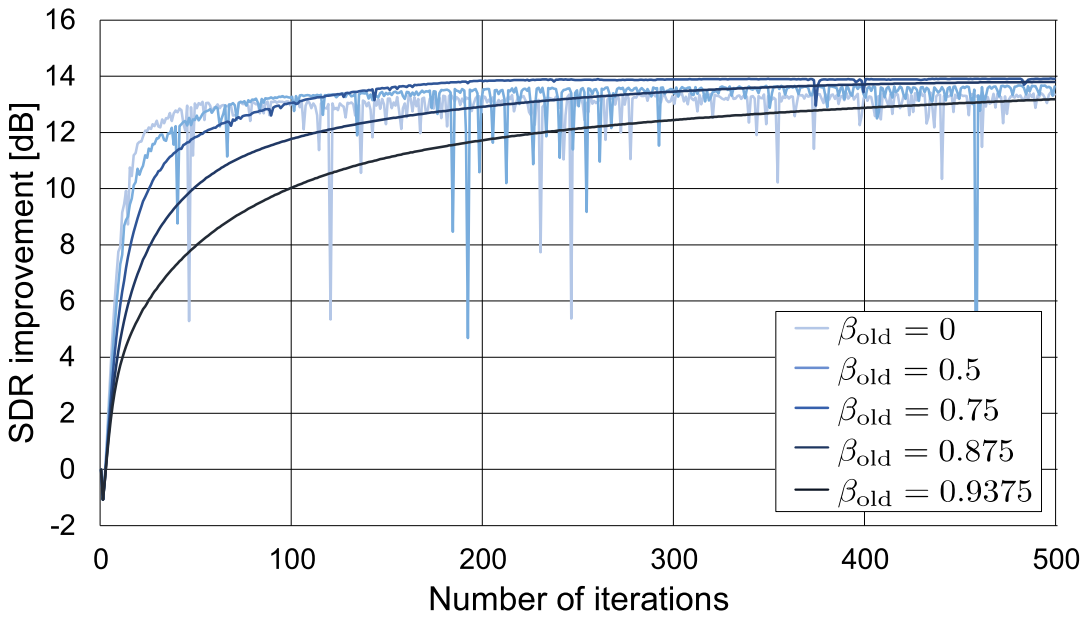

Fig. B.19. Example of convergence behaviors of the MHPSS-based proposed method with various β_{old} and β (song no. 20).

付録 C

各楽曲における従来の BSS との性能比較

各楽曲における従来の BSS との性能比較を Figs. C.1–C.20 に示す．Figs. C.1–C.20 は，4.5 節の Fig. 4.7 に示した平均性能に対応する，各楽曲における性能比較である．

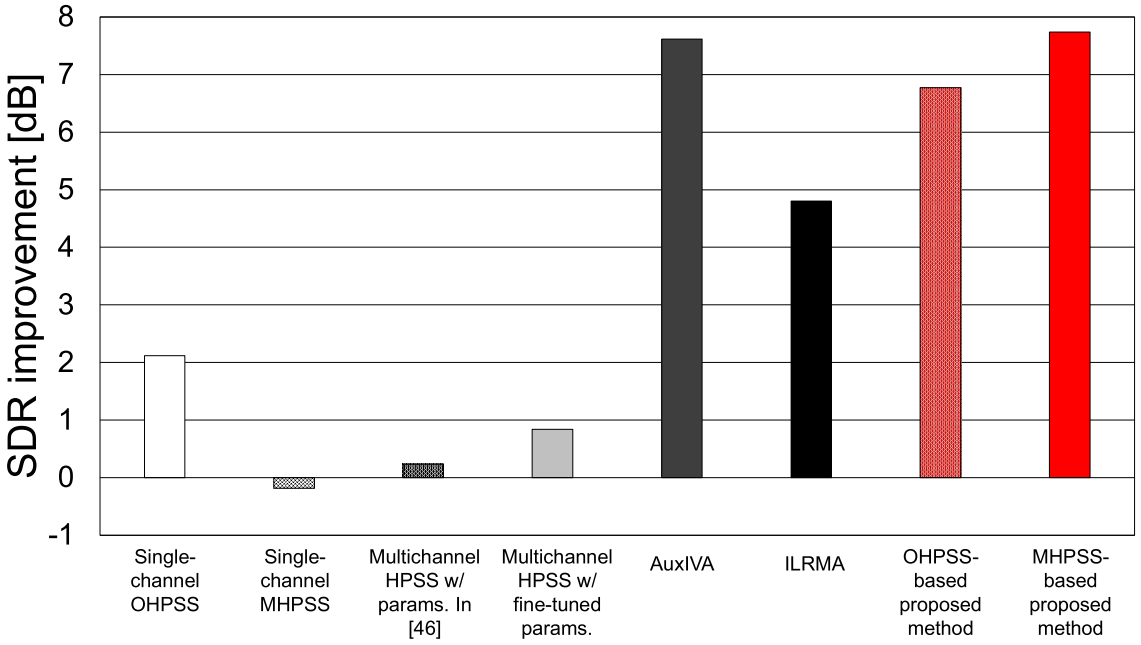

Fig. C.1. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 1).

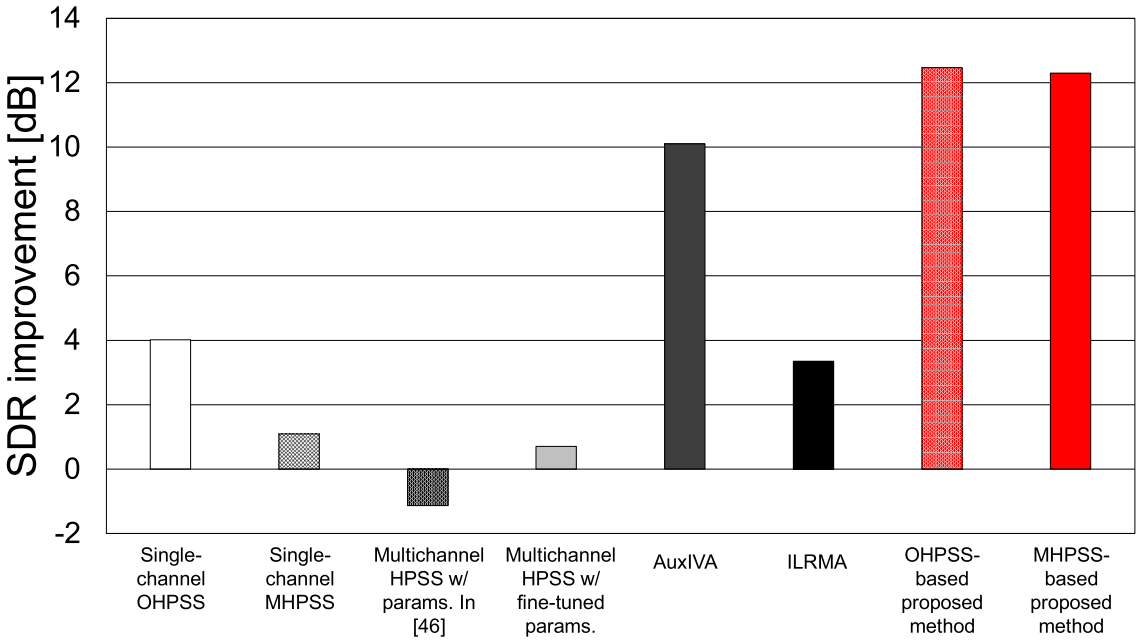

Fig. C.2. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 2).

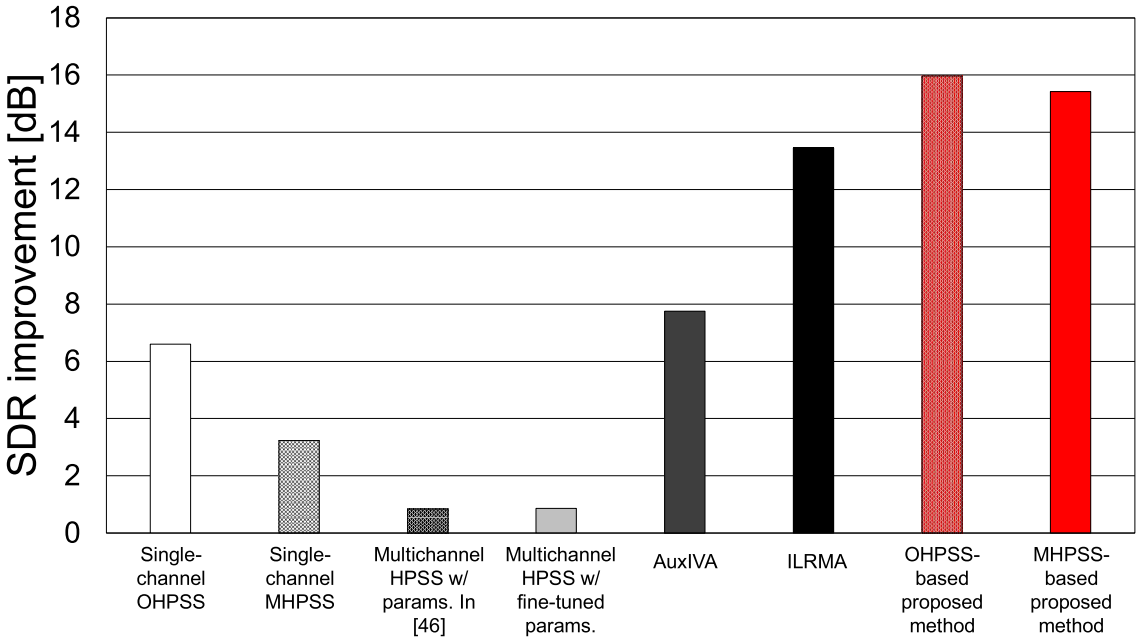

Fig. C.3. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 3).

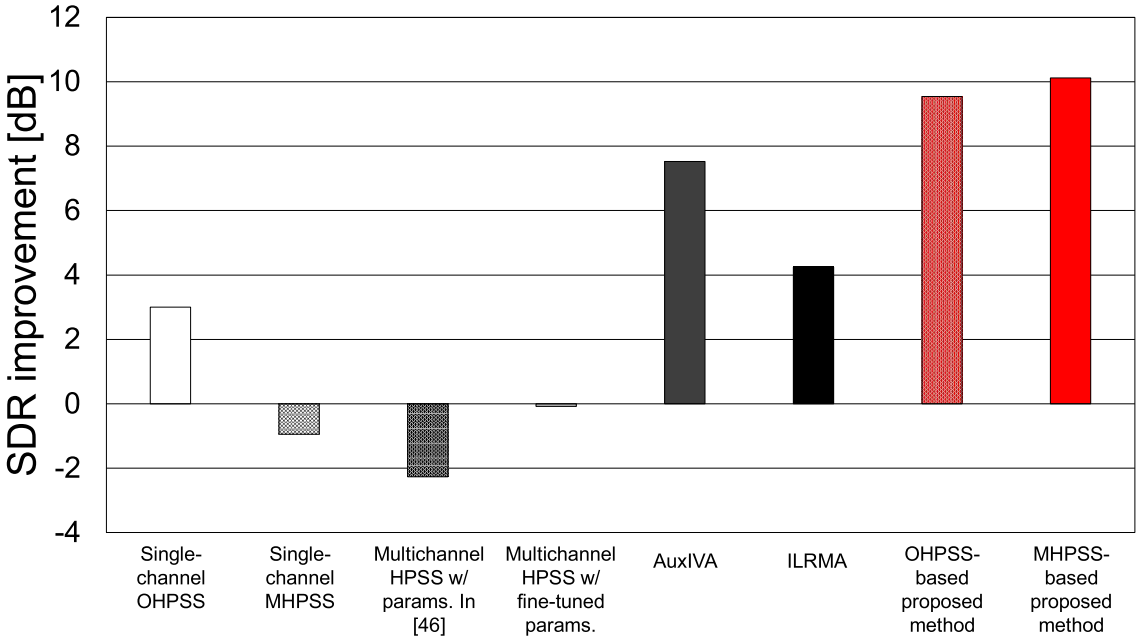

Fig. C.4. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 4).

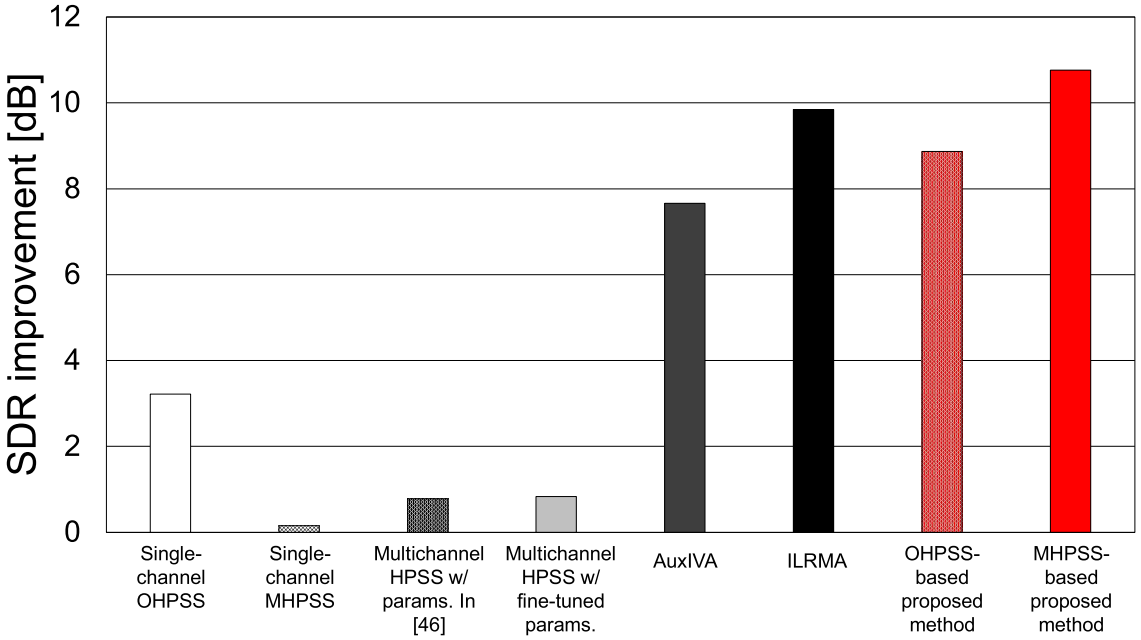

Fig. C.5. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 5).

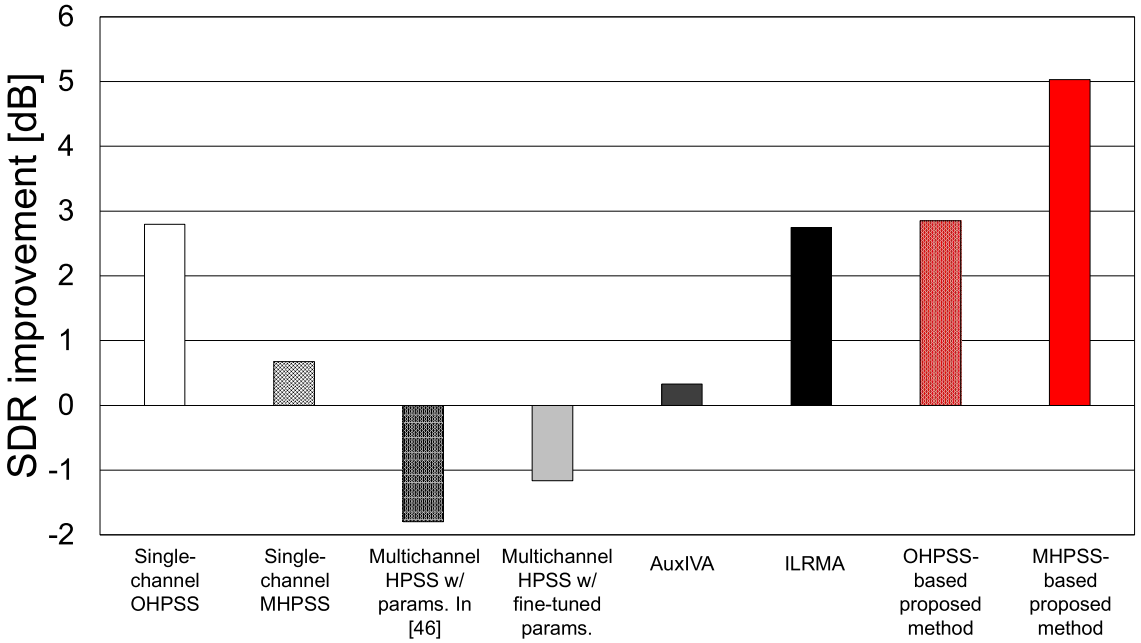

Fig. C.6. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 6).

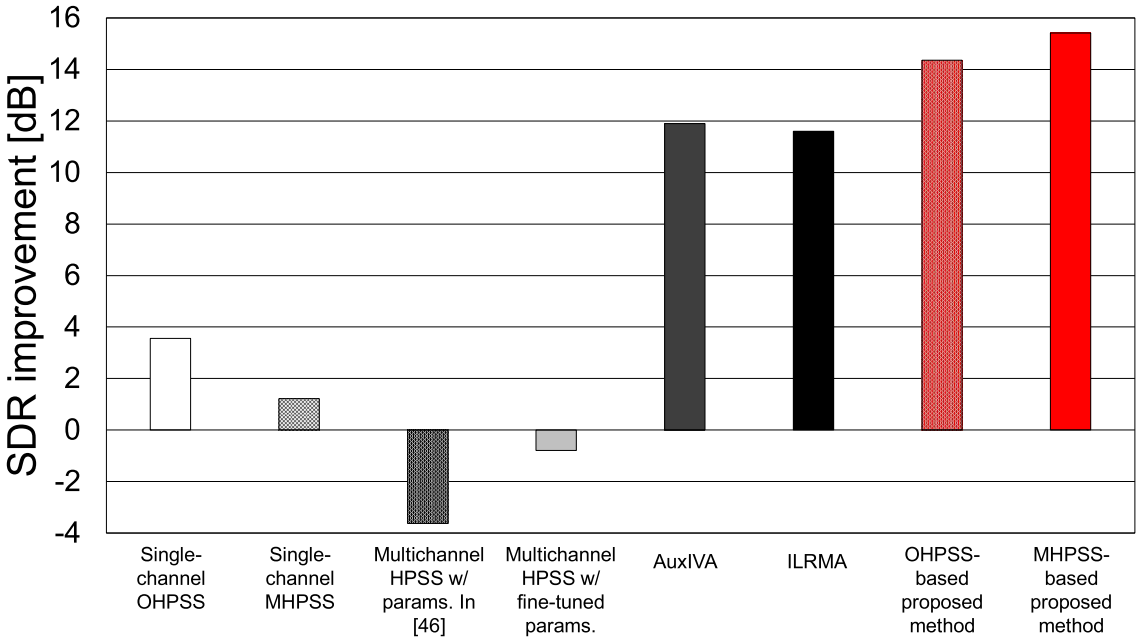

Fig. C.7. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 7).

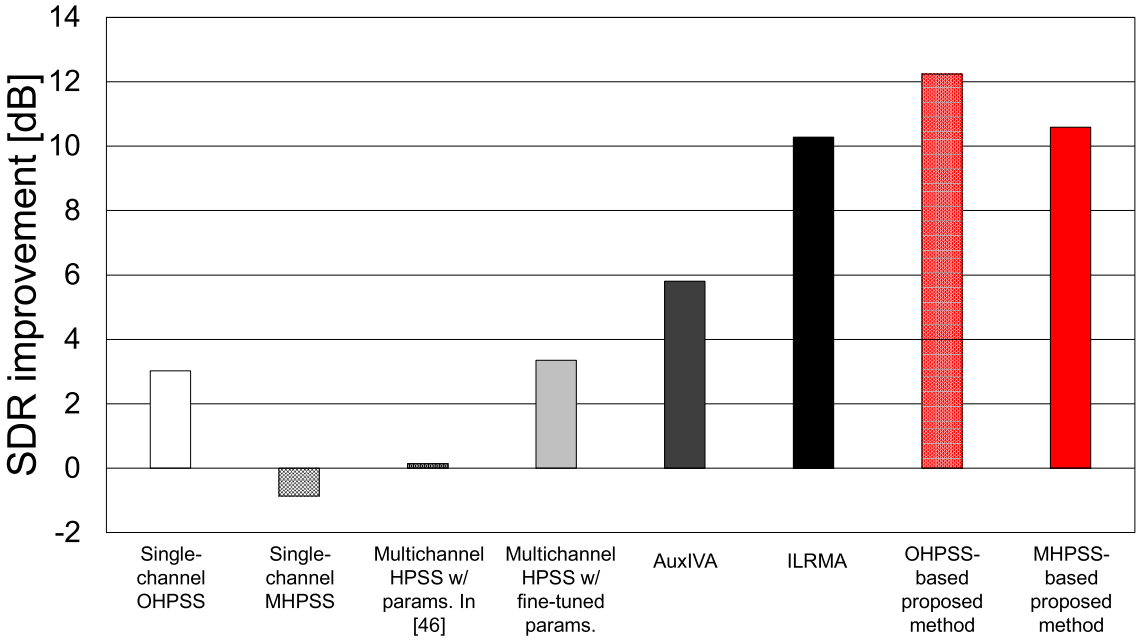

Fig. C.8. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 8).

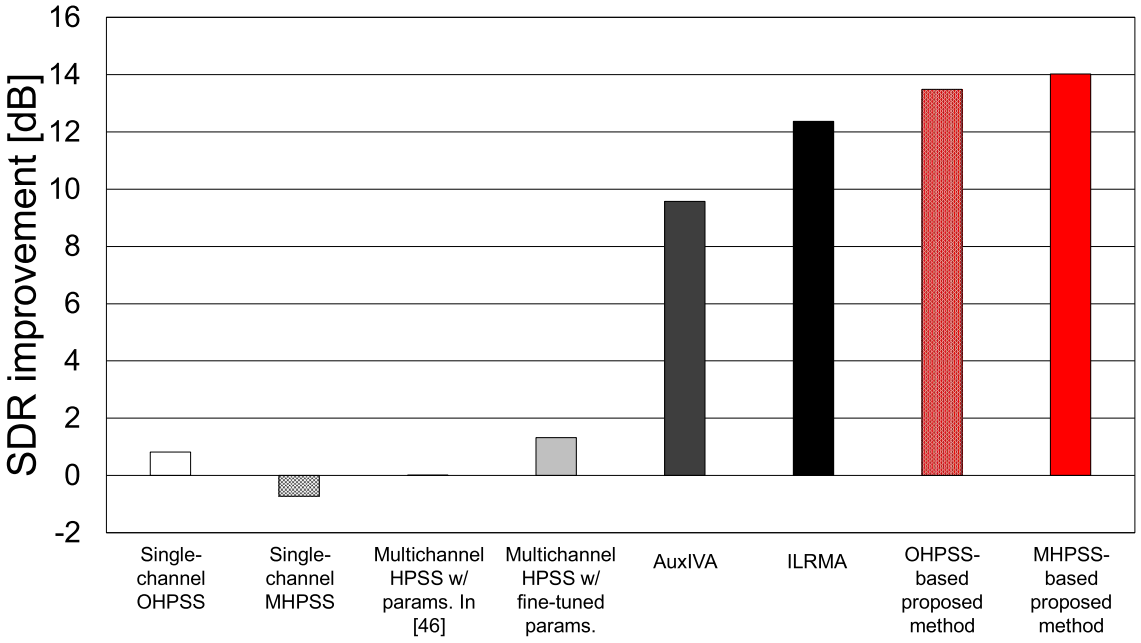

Fig. C.9. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 9).

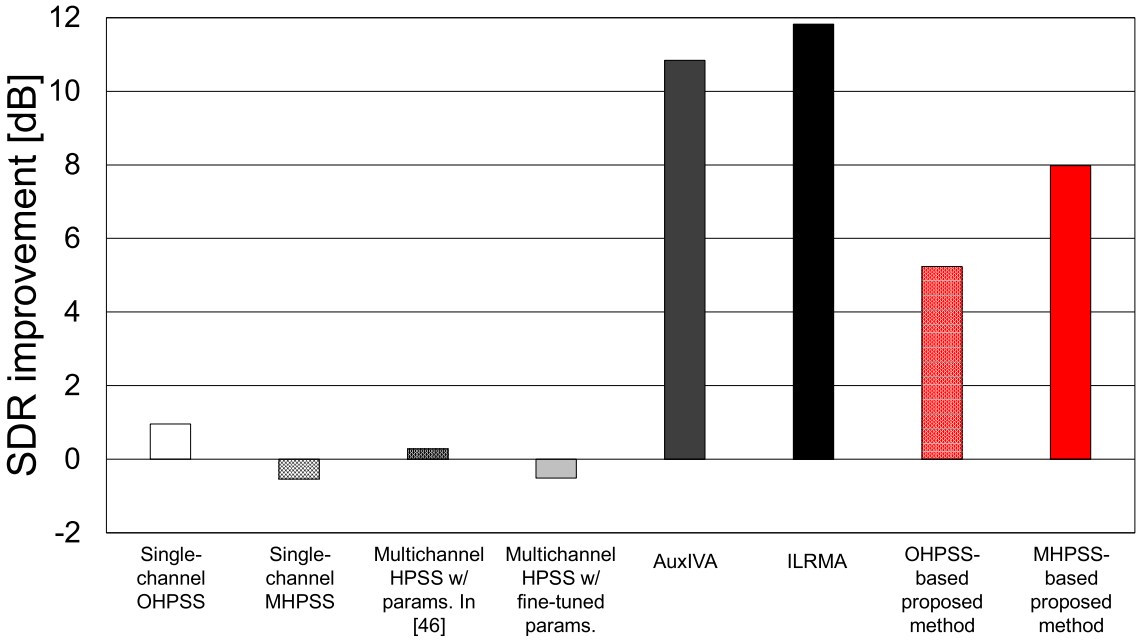

Fig. C.10. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 10).

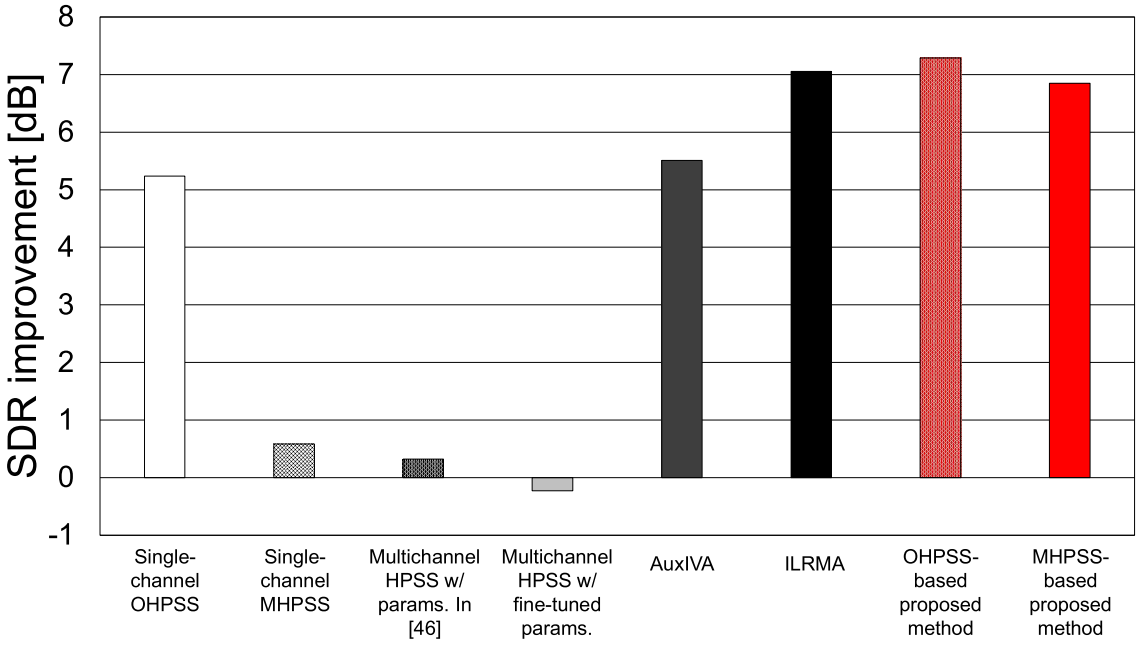

Fig. C.11. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 11).

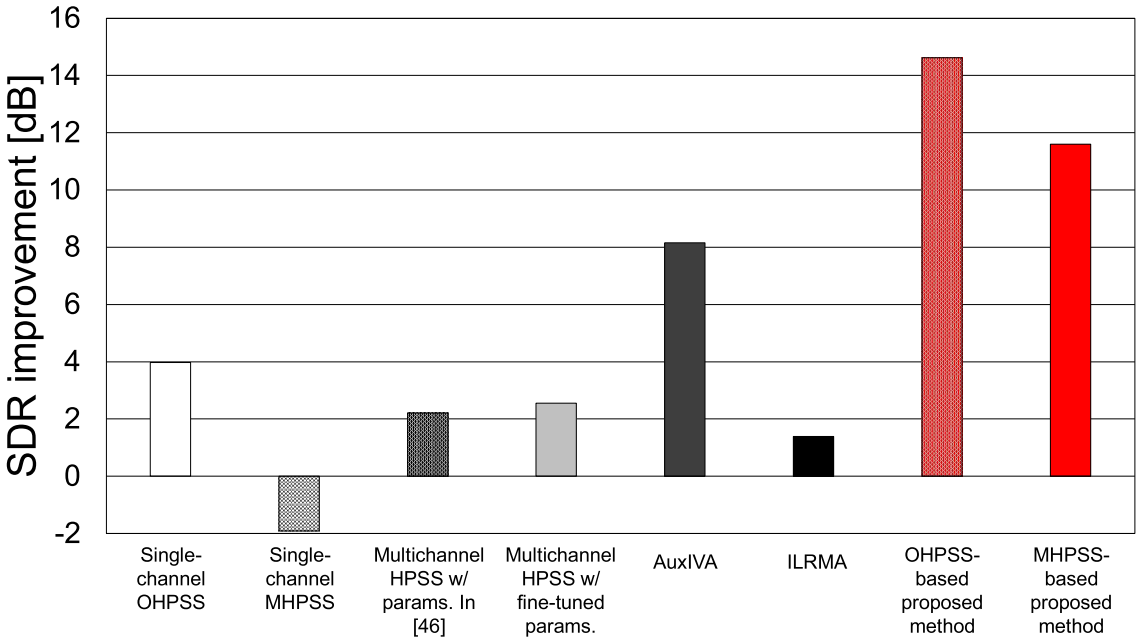

Fig. C.12. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 12).

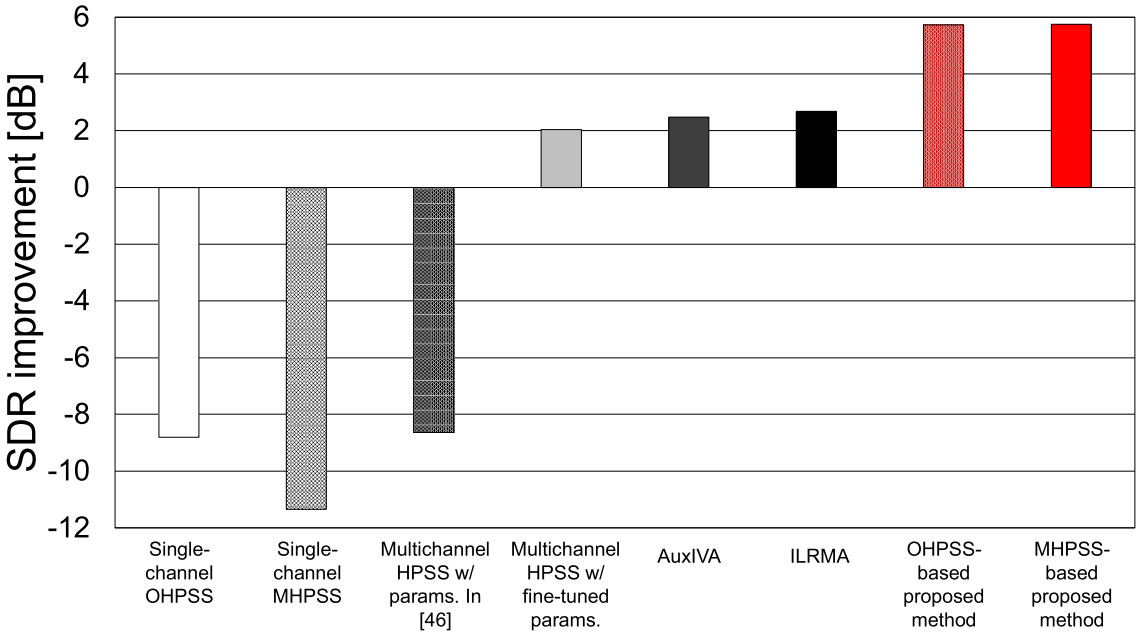

Fig. C.13. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 13).

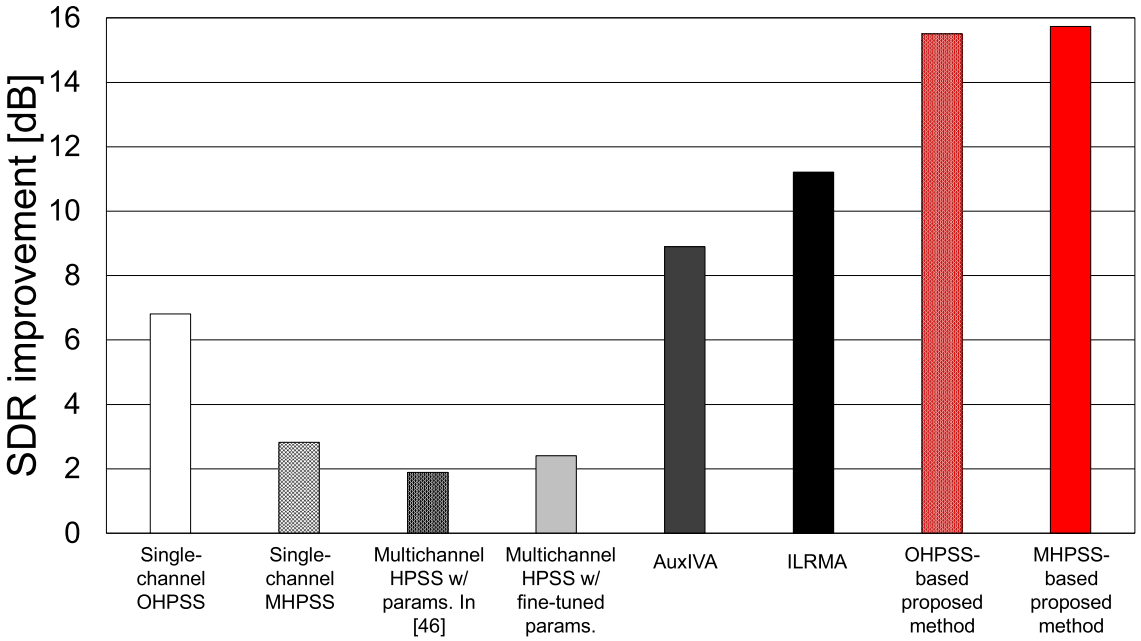

Fig. C.14. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 14).

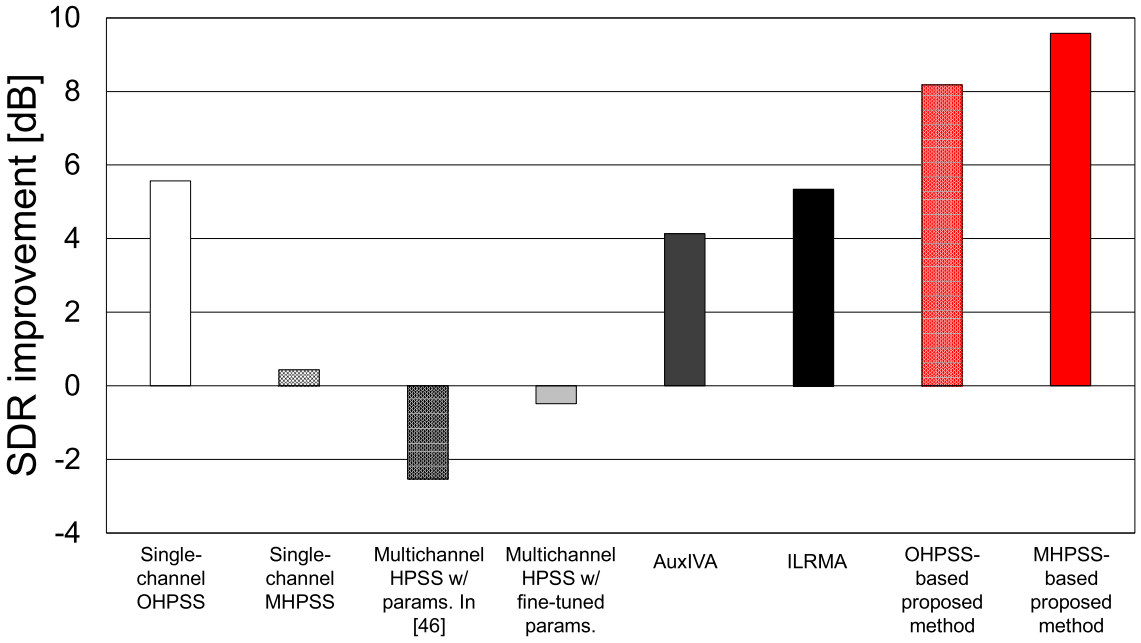

Fig. C.15. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 15).

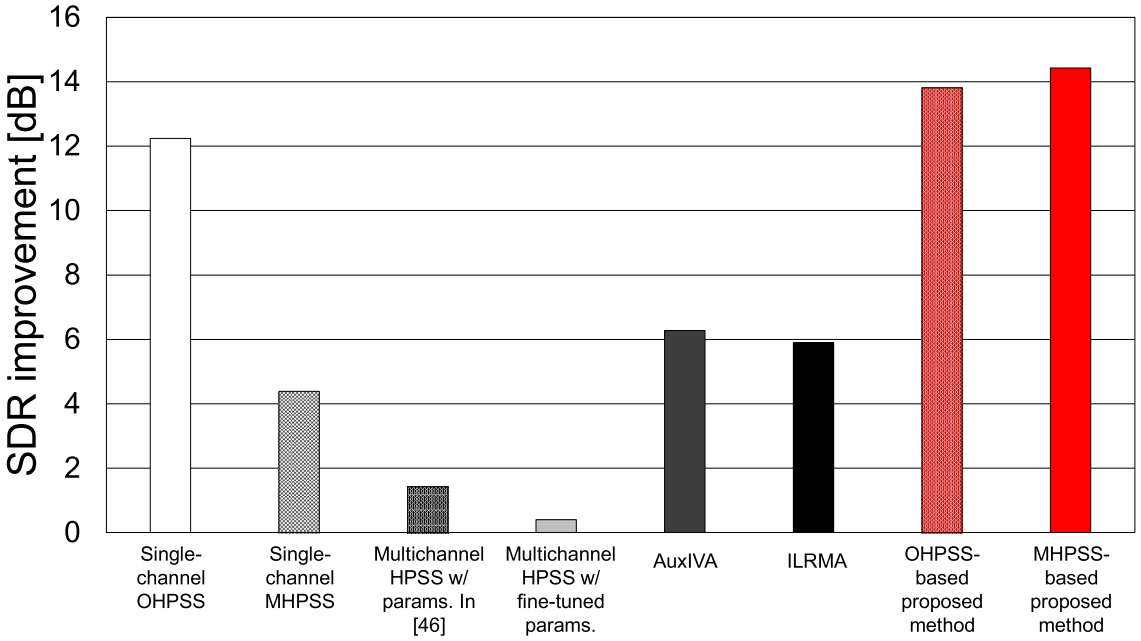

Fig. C.16. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 16).

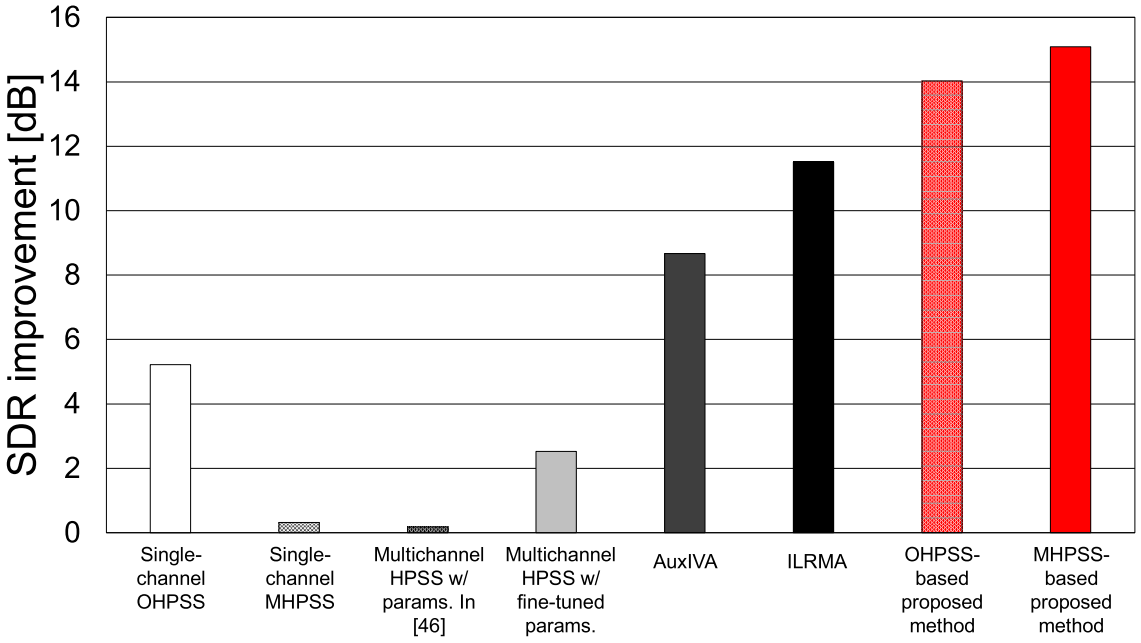

Fig. C.17. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 17).

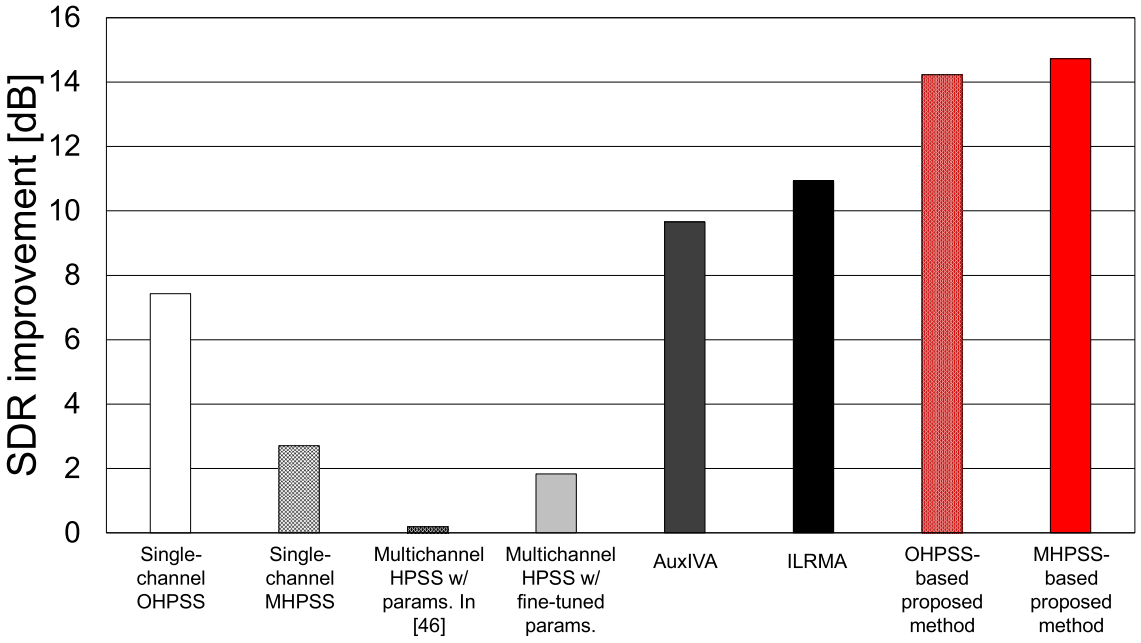

Fig. C.18. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 18).

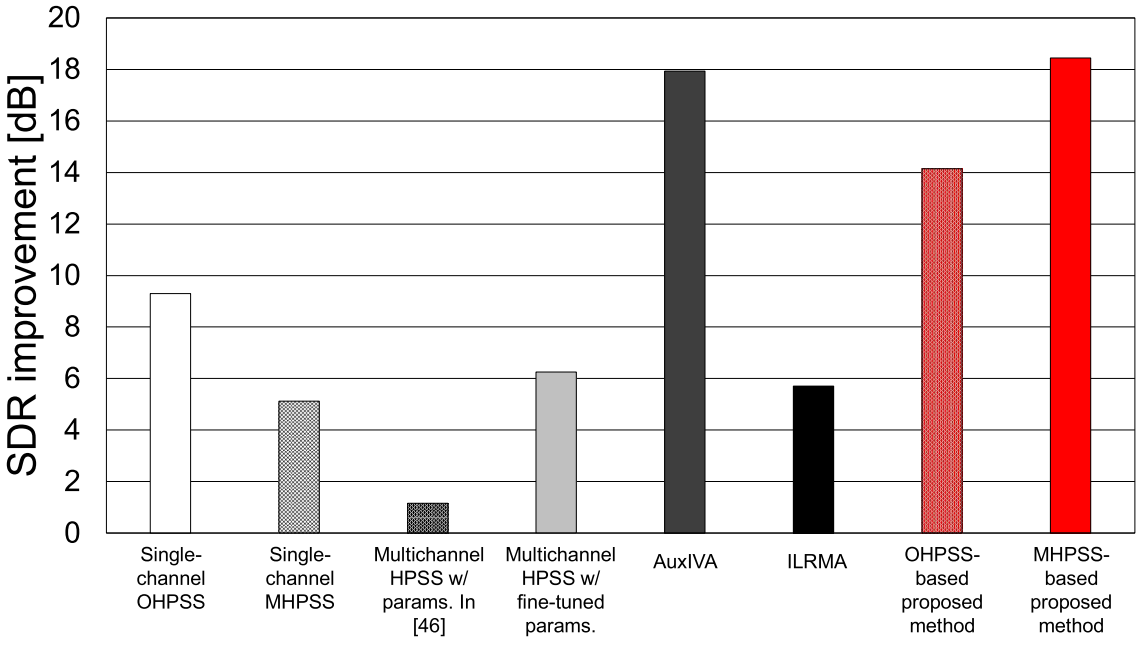

Fig. C.19. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 19).

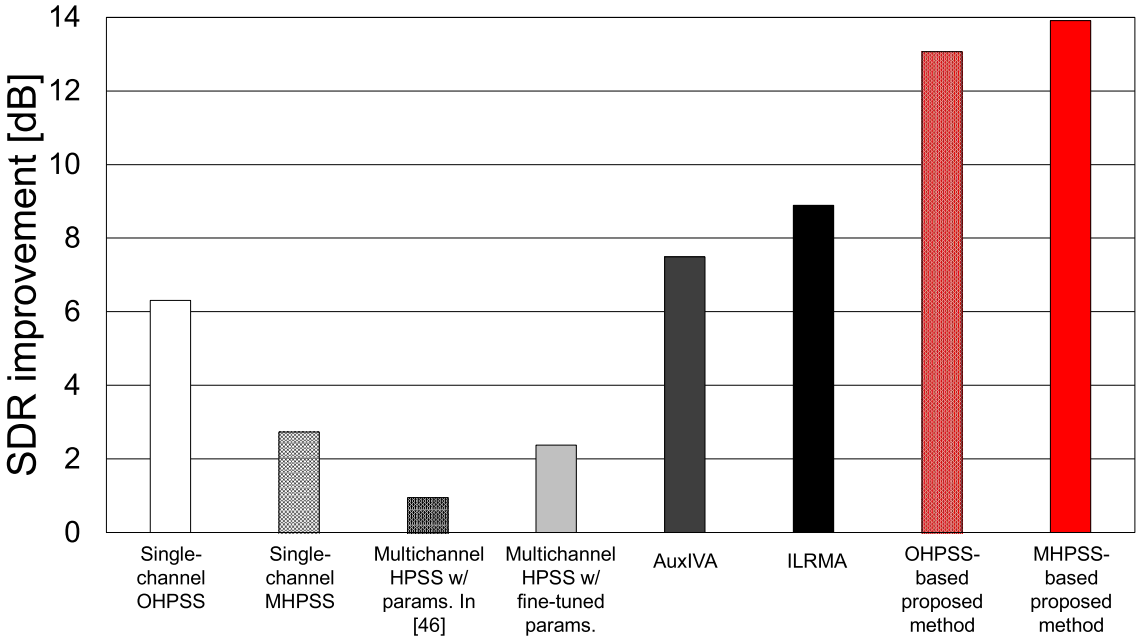

Fig. C.20. Example of SDR improvements of ILRMA, AuxIVA, conventional HPSS, and proposed methods (song no. 20).