

過決定条件 BSS におけるランク 1 空間制約の緩和*

◎北村大地 (総研大), 小野順貴 (NII/総研大), 澤田宏 (NTT), 亀岡弘和 (東大/NTT), 猿渡洋 (東大)

1 はじめに

ブラインド音源分離 (blind source separation: BSS) とは、音源位置や混合系が未知の条件下で観測された信号のみから混合前の元信号を推定する信号処理技術である。過決定条件 (音源数 \leq 観測チャンネル数) における BSS では、独立成分分析 (independent component analysis: ICA) [1] に基づく手法が主流であり、盛んに研究されてきた [2]。一方、モノラル信号等を対象とした劣決定条件 (音源数 $>$ 観測チャンネル数) 下では、非負値行列因子分解 (nonnegative matrix factorization: NMF) [3] を応用した手法が注目を集めている。BSS は一般的に、話者分離や雑音抑圧が目的であるが、音楽を対象とした音源分離の研究も増加している [4]。

劣決定条件にも対応可能な BSS 技術として、従来の NMF を多チャンネル信号用に拡張した多チャンネル NMF (multichannel NMF: MNMF) [5] が提案されている。MNMF は、音源の空間情報に相当するチャンネル間相関を用いて、推定したスペクトル基底を音源毎にクラスタリングすることで分離信号を得る。しかし、MNMF は音源の空間推定とスペクトル推定を同時に行う最適化であり、そのモデルの複雑さから計算コストが極めて高いうえ、初期値依存性が強く、分離精度が安定しない問題がある。

一方、過決定条件における周波数領域 ICA (frequency-domain ICA: FDICA) や ICA の多変量モデルである独立ベクトル分析 (independent vector analysis: IVA) [6] では、時間周波数領域での線形時不変混合を仮定する。しかし、収録環境の残響が長い場合には、混合系を線形時不変モデルで表現できず分離性能が劣化してしまう。著者が近年提案したランク 1 空間モデル制約付き MNMF [7, 8] においても、推定する空間相関行列をランク 1 近似することで線形時不変混合系を仮定しているため、高残響下では同様の問題が生じる。

本稿では、残響等の影響で線形時不変混合仮定が成り立たない場合における分離精度劣化の問題を解決するために、過決定条件における余剰な観測チャンネルを、各音源の残響成分の推定に活用するアルゴリズムを新たに提案する。また、提案するアルゴリズムによって、ランク 1 空間モデル制約付き MNMF の利点である高い計算効率を保ったまま、分離精度の劣化を回避できることを実験的に示す。

2 従来手法

2.1 線形時不変混合仮定

音源数と観測チャンネル数をそれぞれ N, M とし、各時間周波数の多チャンネルの音源信号、観測信号、分離信号をそれぞれ、

$$\mathbf{s}_{ij} = (s_{ij,1} \cdots s_{ij,N})^T \quad (1)$$

$$\mathbf{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^T \quad (2)$$

$$\mathbf{y}_{ij} = (y_{ij,1} \cdots y_{ij,N})^T \quad (3)$$

と表す (要素はすべて複素数)。ここで、 $i=1, \dots, I$ は周波数インデックス、 $j=1, \dots, J$ は時間インデック

ス、 $n=1, \dots, N$ は音源インデックス、 $m=1, \dots, M$ はチャンネルインデックスを示し、 \top は転置を表す。

混合系が線形時不変と仮定すると、各時間フレームにおいて周波数毎の複素混合行列 $A_i = (\mathbf{a}_{i,1} \cdots \mathbf{a}_{i,N})$ ($\mathbf{a}_{i,n}$ は各音源のステアリングベクトル) を定義でき、観測信号を次式で表現できる。

$$\mathbf{x}_{ij} = A_i \mathbf{s}_{ij} \quad (4)$$

この混合系を Fig. 1 (a) に示す。線形時不変混合系では、全ての時間フレームが他の時間フレームと独立し、互いに影響を及ぼさないことを意味している。しかし残響が多い場合は、Fig. 1 (b) のように前の時間フレームの残響成分が現在の時間フレームに漏れ出すため、 A_i だけでは表現できなくなる。従って、このような線形時不変混合仮定は、各音源から各マイクロフォンまでのインパルス応答が、短時間フーリエ変換 (short-time Fourier transform: STFT) の窓関数と比べて十分に短い場合に成立する。

線形時不変混合系において $M=N$ とすれば、分離ベクトル $\mathbf{w}_{i,n}$ で表現される分離行列 $W_i = (\mathbf{w}_{i,1} \cdots \mathbf{w}_{i,N})^H$ が存在し、分離信号を次式で表現できる。

$$\mathbf{y}_{ij} = W_i \mathbf{x}_{ij} \quad (5)$$

但し、 H はエルミート転置を表す。

2.2 事前処理に主成分分析を用いる BSS

従来の FDICA や IVA による音源分離では、過決定条件 $N < M$ の場合に、事前処理として主成分分析 (principle component analysis: PCA) による次元圧縮を行い、 $N=M$ とすることが一般的である。これは、観測信号中の残響成分が PCA によって主成分に射影されることを期待しており、多少の残響が存在する場合においても前述の線形時不変混合仮定が成り立つようにする為である。しかしながら、音楽信号や高雑音下での話者分離等のように、各音源の混合パワーが著しく偏っている場合には、パワーの弱い音源の主成分が PCA によって除かれてしまう危険がある。また、残響が強い場合は、PCA を施しても十分な残響抑圧の効果が得られず、線形時不変混合が成立せずに音源分離精度が劣化する。

線形時不変混合仮定は、STFT における窓関数を長くすることで成立しやすくなる。しかし、FDICA や IVA では、長すぎる窓関数を用いると、極めて狭帯域な信号間の独立性を分離尺度に利用することになるため、音源分離が困難になってしまう。従って、窓関数の長さに関しては、線形時不変混合仮定と音源の独立性という二つの観点から、音源分離性能がトレードオフになることが知られている。

2.3 ランク 1 空間モデル制約付き MNMF

MNMF では、観測信号は次式のようにチャンネル間相関行列 X_{ij} で表現される [5]。

$$X_{ij} = \mathbf{x}_{ij} \mathbf{x}_{ij}^H \quad (6)$$

*Relaxation of rank-1 spatial model in overdetermined BSS by Daichi Kitamura (SOKENDAI), Nobutaka Ono (NII/SOKENDAI), Hiroshi Sawada (NTT), Hirokazu Kameoka (The University of Tokyo/NTT), Hiroshi Saruwatari (The University of Tokyo)

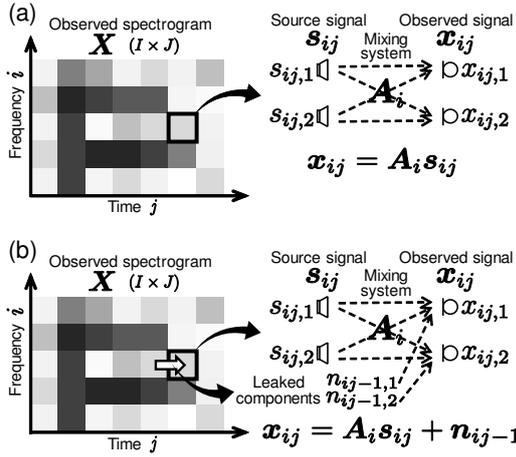


Fig. 1 Mixing system of each spectrogram slot when $N=M=2$; (a) holds linear time-invariant mixing system and there is no reverberation, (b) has some leaked components from previous frame because of reverberation.

この \mathbf{X}_{ij} を近似する MNMF の分解モデル $\hat{\mathbf{X}}_{ij}$ は次式で定義される.

$$\mathbf{X}_{ij} \approx \hat{\mathbf{X}}_{ij} = \sum_k (\sum_n \mathbf{H}_{i,n} z_{nk}) t_{ik} v_{kj} \quad (7)$$

ここで, $k=1, \dots, K$ は NMF における基底 (スペクトルパターン) のインデックスを示し, $\mathbf{H}_{i,n}$ は周波数 i における音源 n の空間相関行列を表す $M \times M$ のエルミート半正定値行列である. また, $z_{nk} \in \mathbb{R}_{\geq 0}$ は k 番目の基底を n 番目の音源に対応付ける潜在変数に相当し, $\sum_n z_{nk} = 1$ であり, $z_{nk} = 1$ のとき, k 番目の基底は n 番目の音源のみに寄与する. さらに, $t_{ik} \in \mathbb{R}_{\geq 0}$ 及び $v_{kj} \in \mathbb{R}_{\geq 0}$ はそれぞれ単一チャンネル NMF の基底行列 \mathbf{T} 及びアクティベーション行列 \mathbf{V} の要素と等価である. MNMF のモデルの概念を Fig. 2 に示す. 劣決定条件においては, Fig. 2 に示す分離行列 \mathbf{W}_i は求まらないが, MNMF では各音源に一意に対応する空間相関行列 \mathbf{H} と全音源のスペクトル成分 \mathbf{TV} を潜在変数 \mathbf{z} でクラスタリングすることで, 分離信号 \mathbf{y} を得る. \mathbf{X}_{ij} と $\hat{\mathbf{X}}_{ij}$ 間の板倉斎藤擬距離は

$$Q_{\text{MNMF}} = \sum_{i,j} \left[\text{tr}(\mathbf{X}_{ij} \hat{\mathbf{X}}_{ij}^{-1}) + \log \det \hat{\mathbf{X}}_{ij} \right] \quad (8)$$

で表され, これを最小化する変数 $\mathbf{H}, \mathbf{Z}, \mathbf{T}$ 及び \mathbf{V} を求める問題となる. しかしながら, この最適化は極めて高い計算コストを必要とし, 分離精度は各変数の初期値に強く依存する問題がある.

式 (8) の効率的な最適化手法として, $\mathbf{H}_{i,n}$ がランク 1 となる制約条件を導入したモデル [7, 8] では, 過決定条件の線形時不変混合を仮定することで, IVA の高速な最適化更新式 [9] と単一チャンネル NMF の最適化更新式の交互反復で全変数の最適化が可能となる. $N=M$ のとき, IVA の更新式は次式となる [9].

$$r_{i,j,n} = \sum_k t_{ik} v_{kj} \quad (9)$$

$$V_{i,n} = \frac{1}{J} \sum_j \frac{1}{r_{i,j,n}} \mathbf{x}_{ij} \mathbf{x}_{ij}^H \quad (10)$$

$$\mathbf{w}_{i,n} \leftarrow (\mathbf{W}_i V_{i,n})^{-1} \mathbf{e}_n \quad (11)$$

$$\mathbf{w}_{i,n} \leftarrow \mathbf{w}_{i,n} (\mathbf{w}_{i,n}^H V_{i,n} \mathbf{w}_{i,n})^{-\frac{1}{2}} \quad (12)$$

$$y_{i,j,n} = \mathbf{w}_{i,n}^H \mathbf{x}_{ij} \quad (13)$$

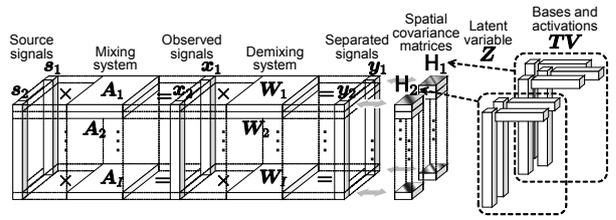


Fig. 2 Conceptual model of MNMF ($N=M=2$).

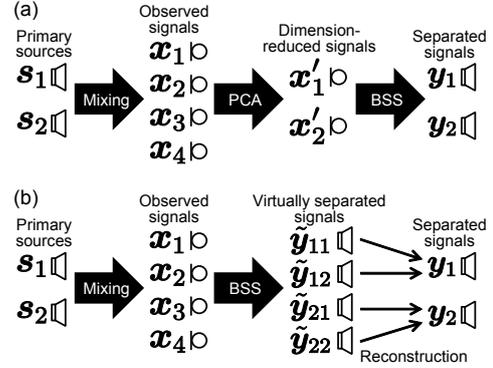


Fig. 3 Algorithms of (a) conventional and (b) proposed methods ($N=2, M=4, P=2$).

但し, \mathbf{e}_n は n 番目の要素のみが 1 の単位ベクトルを示す. さらに, NMF の更新式は次式で与えられる.

$$t_{il,n} \leftarrow t_{il,n} \sqrt{\frac{\sum_j |y_{ij,n}|^2 v_{lj,n} (\sum_{l'} t_{il',n} v_{l',n})^{-2}}{\sum_j v_{lj,n} (\sum_{l'} t_{il',n} v_{l',n})^{-1}}} \quad (14)$$

$$v_{lj,n} \leftarrow v_{lj,n} \sqrt{\frac{\sum_i |y_{ij,n}|^2 t_{il,n} (\sum_{l'} t_{il',n} v_{l',n})^{-2}}{\sum_i t_{il,n} (\sum_{l'} t_{il',n} v_{l',n})^{-1}}} \quad (15)$$

ここで, $l=1, \dots, L$ はある一つの音源に関する基底のインデックスであり, $t_{il,n}$ 及び $v_{lj,n}$ は音源 n を表現する基底とアクティベーションである. 式 (8) のように, 潜在変数を用いて全 K 本の基底を各音源に適応的に割り当てるモデルへの拡張も可能である [7, 8].

ランク 1 空間モデル制約付き MNMF では, 非常に高速に全変数 \mathbf{W}_i, \mathbf{T} 及び \mathbf{V} を最適化でき, 制約無し MNMF と同程度の分離性能を達成することが可能である. しかし, 残響の影響が強くなると, ランク 1 空間モデルの近似が成り立たなくなるため, 分離精度は著しく劣化する. 制約無し MNMF では, フルランクの $\mathbf{H}_{i,n}$ の推定が成功すれば, 多少の残響が存在していても比較的高い精度で分離できる.

3 提案手法

3.1 過決定条件における余剰観測チャンネルを用いたランク 1 制約の緩和

従来手法のランク 1 空間モデル制約 (線形時不変混合仮定) を緩和するために, 過決定条件における余剰な観測チャンネルを残響成分のモデル化に活用する手法を新たに提案する. 今, N 個の音源の P 倍の観測チャンネル $M(=PN)$ が得られる場合を想定する. 従来の過決定条件 BSS では, Fig. 3 (a) に示すように, PCA による次元圧縮を行い $M=N$ とするが, 提案手法では Fig. 3 (b) のように, 仮想的に M 個の分離信号を推定

する。提案手法での推定音源を次式で表す。

$$\tilde{\mathbf{y}}_{ij} = (\tilde{y}_{ij,11} \cdots \tilde{y}_{ij,1P} \tilde{y}_{ij,21} \cdots \tilde{y}_{ij,2P} \cdots \tilde{y}_{ij,NP})^T \quad (16)$$

$$y_{ij,n} = \sum_p \tilde{y}_{ij,np} \quad (17)$$

仮想的な分離信号 $\tilde{\mathbf{y}}_{ij}$ は、各音源の直接音に加え、Fig. 1 (b) に示すような、前の時間フレームから漏れる残響成分が新たな別の音源として推定されることを期待している。しかし、BSS 後に出力される信号の順番は一意に定まらないため、直接音成分と残響成分をクラスタリングして音源毎にまとめる必要がある。最終的に、クラスタリングされた信号を式 (17) のように足し合わせることで最終的な分離信号を得る。

3.2 スペクトルの相互相関によるクラスタリング

仮想分離信号 $\tilde{\mathbf{y}}_{ij}$ の各信号が、どの音源の直接音成分及び残響成分かを定めるために、各信号のパワースペクトルの時間フレーム遅れを考慮した相関を全信号間で計算し、相関が高いものから順にマージしてゆくクラスタリングを用いる (Fig. 4 参照)。但し、処理後のクラスタ数は音源数 N であり、要素の少ないクラスタを優先してマージする特殊な階層的クラスタリングを行う。時間フレーム遅れを考慮した相関値 C は下記のように計算される。

$$C(\mathbf{A} \parallel \mathbf{B}) = \max(\{ \sum_{i,j} a_{ij} b_{ij+\tau} | \tau = 0, 1, 2 \}) \quad (18)$$

ここで、 $\mathbf{A} (\in \mathbb{R}_{\geq 0}^{I \times J})$ 及び $\mathbf{B} (\in \mathbb{R}_{\geq 0}^{I \times J})$ はパワースペクトログラムであり、 a_{ij}, b_{ij} はそれぞれ \mathbf{A}, \mathbf{B} の要素を示す。 τ は考慮する時間フレーム遅れであり、遅れ無し ($\tau=0$) から $\tau=2$ フレームまでの遅れを考慮する。このように時間フレーム遅れを考慮する理由は、ある音源の残響成分が実際に直接音成分から 1 フレーム程度遅れて生じるためであり、これを考慮しなければクラスタリングが失敗する危険がある。

3.3 基底共有 MNMF によるクラスタリング

ランク 1 空間モデル制約付き MNMF に、余剰チャンネルによる制約緩和を適用する場合、同じ音源の直接音成分と残響成分を一つの基底セットで表現する制約を導入することで、音源のクラスタリングを行いつつ分離信号を推定することができる。即ち、ある成分 $\tilde{y}_{ij,n1} \cdots \tilde{y}_{ij,nP}$ を同じ基底セット $t_{i1,n} \cdots t_{iL,n}$ で表現することで、同じ音源の直接音成分と残響成分が推定される仕組みである。このとき、アクティベーションは共有せず $v_{l,j,np}$ として個別に与えることで、直接音成分と残響成分は、同じスペクトルで異なる時間変化を持つという音源にモデル化できる。この基底共有を導入した場合のランク 1 空間モデル制約付き MNMF のコスト関数は下記のようになる。

$$Q = \sum_{i,j} \left[\sum_{n,p} \frac{|\tilde{y}_{ij,np}|^2}{\sum_l t_{il,n} v_{l,j,np}} - 2 \log |\det \mathbf{W}_i| + \sum_{n,p} \log \sum_l t_{il,n} v_{l,j,np} \right] \quad (19)$$

式 (19) を最小化する \mathbf{W}_i の更新式は式 (9)–(13) において $N \leftarrow M = NP$ とした場合と同様であり、NMF 変数の更新式は下記のようになる。

$$t_{il,n} \leftarrow t_{il,n} \sqrt{\frac{\sum_{j,p} |\tilde{y}_{ij,np}|^2 v_{l,j,np} (\sum_{l'} t_{il',n} v_{l',j,np})^{-2}}{\sum_{j,p} v_{l,j,np} (\sum_{l'} t_{il',n} v_{l',j,np})^{-1}}} \quad (20)$$

$$v_{l,j,np} \leftarrow v_{l,j,np} \sqrt{\frac{\sum_i |\tilde{y}_{ij,np}|^2 t_{il,n} (\sum_{l'} t_{il',n} v_{l',j,np})^{-2}}{\sum_i t_{il,n} (\sum_{l'} t_{il',n} v_{l',j,np})^{-1}}} \quad (21)$$

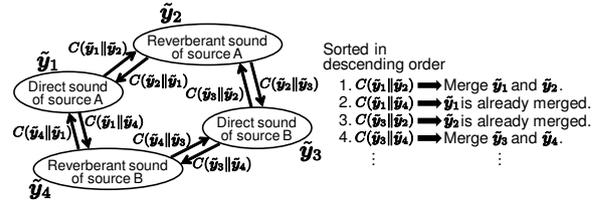


Fig. 4 Atypical hierarchical clustering using correlation C ($N=2, M=4, P=2$).

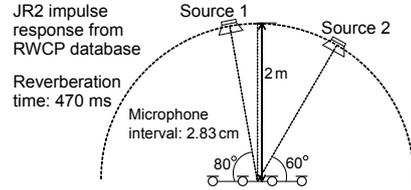


Fig. 5 Recording condition of room impulse response.

しかし、基底数 L や初期値によっては、共有基底が一つの音源を表現せずに、複数音源の直接音成分のみ、あるいは複数音源の残響成分のみを表現してしまう場合がある。これを避けるため、本稿では $M=NP$ チャンネルの IVA (PCA を事前に用いない IVA) の推定分離行列 \mathbf{W}_i を先に求め、前述の階層的クラスタリングによって分離ベクトルを正しい順序 (一つの基底セットと同じ音源の直接音成分及び残響成分が対応する順序) に並び替えた \mathbf{W}_i を初期値として与える。

4 評価実験

4.1 実験条件

提案手法の有効性を確認するために、音楽信号を対象とした分離評価実験を行った。実験では、音源数 $N=2$ 、観測チャンネル数 $M=4$ の過決定条件の観測信号を作成するために、RWCP [10] に収録されているインパルス応答 (JR2, Fig. 5 参照) を各音源信号に畳み込んだ。音源信号は Table 1 に示すように、SiSEC [11] の 3 種の音楽データ、各 2 楽器を選択した。比較手法は、PCA を事前処理に用いる IVA (PCA+IVA)、PCA を事前処理に用いるランク 1 空間モデル制約付き MNMF (PCA+Rank1 MNMF)、ランク 1 制約緩和を行い 3.2 節の手法で信号を再構成する IVA (Proposed IVA)、ランク 1 制約緩和を行い式 (20), (21) を用いる基底共有型ランク 1 空間モデル制約付き MNMF (Proposed Rank1 MNMF) である。但し、Proposed Rank1 MNMF の \mathbf{W}_i の初期値は、Proposed IVA で推定された $w_{i,m}$ を正しい順序に並び替えたものを与える。さらに、従来の制約無し MNMF [5] も比較対象に含める。推定された空間相関行列から SN 比最大化ビームフォーマ [12] を構成する線形時不変フィルタとしての分離 (MNMF w/o MWF)、多チャンネル Wiener フィルタを適用する手法 (MNMF+MWF) の 2 つを比較する。最後に参考値として、各音源の真の空間相関行列の時間平均を用いた理想的な SN 比最大化ビームフォーマ (Optimal linear filter) の精度とも比較する。MNMF+MWF 以外の手法は全て projection back [13] をかけ、信号を正しいスケールに戻す必要がある。その他の実験条件は Table 2 に示す。Figure 5 に示す 470 ms のインパルス応答に対し、128 ms の解析窓を用いており、ランク 1 空間モデルが成立しない条件である。分離精度を示す客観評価値には signal-to-distortion ratio (SDR) [14] を用いた。SDR は、非目的音の除去性能と人工歪みの少なさを含む総合的な分離性能である。

Table 1 Music sources

ID	Song	Source (1/2)
1	bearlin-roads_snip_85_99	acoustic_guit_main/piano
2	fort_minor-remember_the_name_snip_54_78	drums/vocals
3	ultimate_nz_tour_snip_43_61	guitar/vocals

Table 2 Experimental conditions

Sampling frequency	Down sampled from 44.1 kHz to 16 kHz
FFT length	128 ms
Window shift	64 ms
Number of bases	$L = 15$ ($K = 30$)
Number of iterations	200

4.2 実験結果

Figure 6 は、各手法において各変数の初期乱数を変えて 10 回試行した際の平均と標準偏差を楽曲毎に示している。いずれの楽曲に対しても、PCA を事前処理に用いる従来手法は低い精度となっており、ランク 1 空間モデル (線形時不変混合仮定) を用いた推定があまりできていないことを示している。一方、Proposed Rank1 MNMF は線形分離手法であるにもかかわらず良好な分離を達成しており、提案手法の有効性が確認できる。特に、Fig. 6 (b), (c) においては、Optimal linear filter の値を上回っており、線形時不変混合仮定の限界精度を超える結果となった。この事実からも、提案手法がランク 1 空間モデル制約をうまく緩和し、分離精度を向上させていることがわかる。MNMF w/o MWF は強い初期値依存性と低い分離精度を示しており、制約の無いフルランク MNMF の空間相関行列 H の推定の困難さが伺える。しかし、MNMF の基底による時変音源モデルと非線形ポスト処理により、MNMF+MWF の分離精度は非常に高い結果となった。

Table 3 は ID3 の楽曲に対する各手法の実行時間を示している。計算は Intel Core i7-4790 (3.60 GHz) の CPU が搭載された PC で、MATLAB 8.3 (64 bit) 上で行った。また、Proposed Rank1 MNMF に関しては、初期値の W_i を求める事前処理 (Proposed IVA) を含んだ計算時間を示している。この結果から、提案手法は高い計算効率を保ったまま、MNMF+MWF と同程度の分離精度を達成していることがわかる。

5 おわりに

本稿では、過決定条件 BSS において、線形時不変混合仮定が成立しない場合における分離精度の向上を目標とし、余剰な観測チャンネルを用いて残響成分を別の音源としてモデル化することで、ランク 1 空間モデル制約を緩和する手法を新たに提案した。提案手法は従来の PCA を事前に用いる手法と比較して、より高精度な分離が可能であることが確認された。

謝辞 本研究の一部は JSPS 特別研究員奨励費 26-10796 の助成を受けたものである。

References

- [1] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol.36, no.3, pp.287–314, 1994.
- [2] 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 Trans. ASLP*, vol.14, no.2, pp.666–678, 2006.
- [3] D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," *Proc. Advances in Neural Information Processing Systems*, vol.13, pp.556–562, 2001.
- [4] H. Kameoka, M. Nakano, K. Ochiai, Y. Imoto, K. Kashino and S. Sagayama, "Constrained and regularized variants

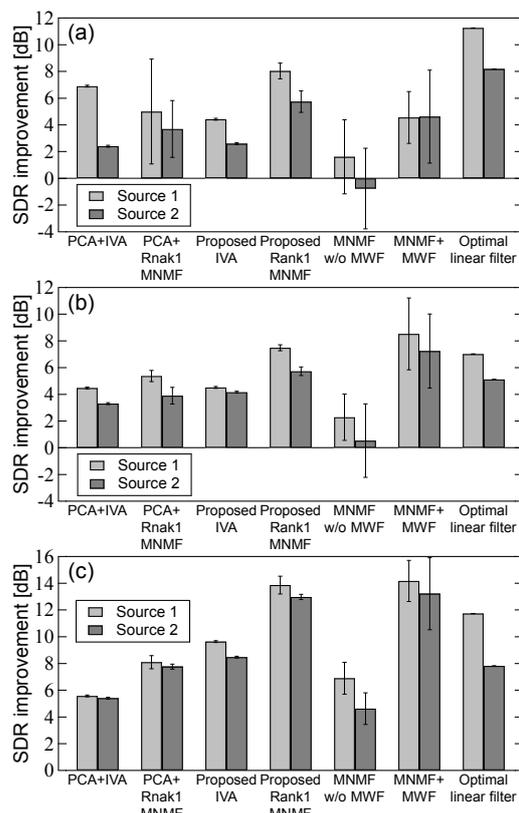


Fig. 6 Average SDR improvements for (a) ID1 song, (b) ID2 song, and (c) ID3 song.

Table 3 Computational times for separation of ID3 (s)

PCA+IVA	PCA+Rnak1 MNMF	Proposed IVA	Proposed Rank1 MNMF	MNMF+MWF
23.4	29.4	60.1	143.9	3611.8

of non-negative matrix factorization incorporating music-specific constraints," *Proc. ICASSP*, pp.5365–5368, 2012.

- [5] H. Sawada, H. Kameoka, S. Araki and N. Ueda, "Multichannel extensions of non-negative matrix factorization with complex-valued data," *IEEE Trans. ASLP*, vol.21, no.5, pp.971–982, 2013.
- [6] T. Kim, H. T. Attias, S.-Y. Lee and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASLP*, vol.15, no.1, pp.70–79, 2007.
- [7] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization with rank-1 spatial model," *Proc. Autumn Meeting of ASJ*, pp.579–582, 2014 (in Japanese).
- [8] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, "Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model," *Proc. ICASSP*, 2015 (in press).
- [9] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *Proc. WASPAA*, pp.189–192, 2011.
- [10] 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," *Proc. LREC*, pp.965–968, 2000.
- [11] S. Araki, F. Nesta, E. Vincent, Z. Koldovsky, G. Nolte, A. Ziehe and A. Benichoux, "The 2011 signal separation evaluation campaign (SiSEC2011):-audio source separation," *Proc. Latent Variable Analysis and Signal Separation*, pp.414–422, 2012.
- [12] H. L. Van Trees, "Detection, Estimation, and Modulation Theory, Optimum Array Processing (Part IV)," *Wiley Interscience*, 2002.
- [13] 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, pp.1–24, 2001.
- [14] E. Vincent, R. Gribonval and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol.14, no.4, pp.1462–1469, 2006.