ランク 1 空間モデル制約付き多チャネル NMF を用いた柔軟索状ロボットにおける雑音抑圧

Noise reduction for a hose-shaped rescue robot using rank-1 multichannel NMF

○高草木萌（筑波大）　北村大地（総研大）
小野順貴（国情研/総研大）　山田武志（筑波大）
牧野昭二（筑波大）　猿渡洋（東京大）

Moe TAKAKUSAKI, University of Tsukuba
Daichi KITAMURA, SOKENDAI
Nobutaka ONO, National Institute of Informatics/ SOKENDAI
Takeshi YAMADA, University of Tsukuba
Shoji MAKINO, University of Tsukuba
Hiroshi SARUWATARI, The University of Tokyo

A hose-shaped rescue robot is one of the robots that are developed for disaster response in case of a large-scale disasters such as a great earthquake. The robot is suitable for entering narrow and dark places covered with rubble in the disaster site, and for finding inside it. This robot can transmit the ambient sound to its operator by using the built-in microphones. However, there is a serious problem that the inherent noise of this robot, such as the vibration sound or the fricative sound, is mixed into the transmitting voice, therefore disturbing the operator's hearing for a call of help from the victim of the disaster. In this paper, we apply the multichannel NMF (nonnegative matrix factorization) with the rank-1 spatial constraint (Rank-1 MNMF), which was proposed by Kitamura *et al.*, to the reduction of the inherent noise.

Key Words: Hose-shaped rescue robot, Inherent noise, Noise reduction, Rank-1 MNMF

1　はじめに

近年多発している大規模災害に対処するためのロボット技術の開発が急務の課題となっている．災害環境下で作業するロボットに求められる役割としては，災害の緊急対応や復旧，人間では困難な作業や危険な作業，そして作業の効率化などが挙げられる．しかし，現在活動している災害対応ロボットは，屋内に比べ屋外での働きが不十分であることや，想定外の事態に対応する能力が低いという課題が残されている．例えば，災害現場で移動できない，災害状況が分からない，失敗すると全体が破たんしてしまう，作業条件が合わない環境では活動できない，などが考えられる．このような従来のロボット技術の問題点を克服する，災害環境作業ロボット技術の開発を支援するため，内閣府総合科学技術・イノベーション会議は，ImPACT 革新的研究開発推進プログラム「タフ・ロボティクス・チャレンジ」[1] を推進している．この研究開発プログラムでは，災害極限状況で効果を発揮するタフな遠隔自律ロボットの実現を目指し，屋外ロボットの基盤となる技術を共同研究開発している．

タフ・ロボティクス・チャレンジでは，五種類の遠隔自律ロボットの開発が検討されている．本稿ではそのうちの一種類である，柔軟索状ロボット [2] (または細径索状ロボット) と呼ばれる蛇のように細長いロボットを扱う．この柔軟索状ロボットに取り付けられたマイクロホンを用いて，災害現場で救助を求める人の声をとらえるための柔軟索状ロボットの音声収録機能の開発を行う．本稿では音声の収録で特に大きな課題となるロボットの内部雑音 (エゴノイズ) の除去を目的とし，ランク 1 空間モデル制約付き多チャネル NMF (Rank-1 multichannel nonnegative matrix factorization: Rank-1 MNMF) [3] による雑音抑圧を柔軟索状ロボットのエゴノイズ除去に適用することを検討する．

2　柔軟索状ロボットのエゴノイズ

2.1　柔軟索状ロボットの仕組みとエゴノイズの性質

本稿で扱う柔軟索状ロボットは，その細長い形状のために，災害現場の瓦礫内の奥深くや細径配管の内部などの狭く暗い場所に進入し，内部を調査することに適している．柔軟索状ロボットの写真を図 1 に，柔軟索状ロボットの構造を図 2 に示す．柔軟索状ロボットは，軸となるホースに繊毛テープを巻き，内蔵された振動モータからの振動で繊毛テープをふるわせることによって，繊毛の向きとは逆の向きにゆっくりと機体が前進するという仕組みになっている．また，オペレータが遠隔操作をするため，ロボットの先端にはカメラと照明が装着されており，さらに慣性計測装置やマイクロホン，スピーカ，ガスセンサなどを内蔵することができる．

柔軟索状ロボットの動作原理により，ロボットの側面にあるマイクロホンに非常に大きな音量の内部雑音（エゴノイズ）が入ってしまう．このエゴノイズの主要な要因は振動モータの振動音や接地面の摩擦音であると考えられる．実際の災害現場では，助けを求める声は聞き取りに十分な音量がなく，このエゴノイズに比べ人の声が小さくなってしまう．このような状況でオペレータが声を聞き取るには，エゴノイズと音声が混ざった収録音から音声のみを分離抽出する必要がある．

2.2　従来のエゴノイズ除去

従来の他のロボットを対象とするエゴノイズ除去の手法は，エゴノイズの音響特性がほとんど変化しないと仮定している．しかし，柔軟索状ロボットにおいてはエゴノイズの特性が接地面によって変化することから，従来手法をそのまま適用しても十分な性能が得られないと考えられる．また，柔軟索状ロボットは災害時に動かすことを想定しているため，事前情報が必要となる手法は適切でない．したがって本稿では，事前情報を必要とせず，従来手法よりも頑健にエゴノイズを除去することができる手法を選

Fig.1 柔軟索状ロボット

Fig.2 柔軟索状ロボットの概図

択し，適用する．

3　ブラインド音源分離のエゴノイズ除去への適用

3.1　アプローチ

事前情報を全く用いずに，観測信号のみから音源を分離する手法は一般にブラインド音源分離と呼ばれ，盛んに研究されている．柔軟索状ロボットには複数のマイクロホンが搭載されているため，音源の空間情報を活用する多チャネル信号を対象としたブラインド音源分離手法が有効利用できると考えられる．そこで，柔軟索状ロボット特有のエゴノイズの分離に有効な多チャネルブラインド音源分離手法について考察する．

エゴノイズの主要な要因は振動モータの振動音や接地面との摩擦音であると考えられる．エゴノイズ信号の時間周波数構造は，数種類の類似するスペクトルの繰り返しになると考えられるため，非負値行列因子分解 (nonnegative matrix factorization: NMF) [4] による表現が有効であると考えられる．また，柔軟索状ロボットは動きが遅いため，エゴノイズの発生源とマイクロホンの位置関係が短時間ではほとんど変化しないという線形時不変混合を仮定すると，線形マイクロホンアレー信号処理による音源分離が効果的である．特に，Kitamura *et al.* が提案した Rank-1 MNMF [3] は，線形マイクロホンアレー信号処理による音源分離手法の一つである独立ベクトル分析 (independent vector analysis: IVA) [5] の音源モデルに，NMF による表現を導入した手法であり，IVA よりも高精度な音源分離を実現している．以上のことから Rank-1 MNMF が柔軟索状ロボットのエゴノイズ除去に有効なのではないかと考え，Rank-1 MNMF を適用する．

3.2　ブラインド音源分離

3.2.1　定式化

まず多チャネルの音源信号，観測信号，分離信号を以下のように示す．音源数とマイクロホン数はともに M とする．

$$\boldsymbol{s}_{ij} = (s_{ij,1} \cdots s_{ij,M})^t \tag{1}$$
$$\boldsymbol{x}_{ij} = (x_{ij,1} \cdots x_{ij,M})^t \tag{2}$$
$$\boldsymbol{y}_{ij} = (y_{ij,1} \cdots y_{ij,M})^t \tag{3}$$

ここで，$1 \leq i \leq I\,(i \in \mathbb{N})$ は周波数インデックスを示し，$1 \leq j \leq J\,(j \in \mathbb{N})$ は時間インデックスを示す．このとき，観測信号は次式のように表される．

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij} \tag{4}$$

ここで，$\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,M})$ は観測信号の混合行列を表している．混合行列と同様に，分離行列も $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,M})^h$ と表すと，分離信号を次式で表すことができる．

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij} \tag{5}$$

ここで，$\boldsymbol{a}_{i,m}$ はステアリングベクトル，$\boldsymbol{w}_{i,m}$ は分離フィルタ，^{h} はエルミート転置を示す．

3.2.2　Rank-1 MNMF

Rank-1 MNMF は，単一チャネルの音源分離に利用される NMF [4] を多チャネル信号も処理できるよう応用した多チャネル NMF (multichannel NMF: MNMF) [6] にランク 1 空間モデル制約を加えて，処理をより効率的にした手法 [3] である．ここでは Kitamura *et al.* によって導出された定式化とアルゴリズム [3] について説明する．観測信号は，MNMF と同じくチャネル間相関行列 X_{ij} で，

$$\mathsf{X}_{ij} = \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^h \tag{6}$$

と表す．この X_{ij} を近似した分解モデル $\hat{\mathsf{X}}_{ij}$ も，MNMF と同じく，

$$\mathsf{X}_{ij} \approx \hat{\mathsf{X}}_{ij} = \Sigma_k (\Sigma_m \mathsf{H}_{i,m} z_{mk}) t_{ik} v_{kj} \tag{7}$$

と定義される．ここで，$m = 1, \cdots, M$ は音源のインデックス，$k = 1, \cdots, K$ は NMF の基底のインデックスを示す．$\mathsf{H}_{i,m}$ は音源 m の周波数 i における $M \times M$ の空間相関行列であり，Rank-1 MNMF では $\mathsf{H}_{i,m} = \boldsymbol{a}_{i,m} \boldsymbol{a}_{i,m}^h$ がランク 1 行列に制約されている．$z_{mk} \in \mathbb{R}_{[\nu,\mu]}$ は，K 本ある NMF 基底 (頻出スペクトル) のそれぞれを各音源に分配する重みであり，$z_{mk} = 1$ のとき k 番目の基底は音源 m にのみ寄与することを示す．また，$t_{ik} \in \mathbb{R}_+$ と $v_{kj} \in \mathbb{R}_+$ は NMF の基底行列 $\boldsymbol{T}$ とアクティベーション行列 $\boldsymbol{V}$ の要素である．MNMF は，空間相関行列 H と音源情報の $\boldsymbol{TV}$ を分配関数である $\boldsymbol{z}$ で割り当てることによって分離信号 $\boldsymbol{y}$ を得るが，Rank-1 MNMF では上記の分解モデルからさらに分離行列 $\boldsymbol{W}_i$ を求めることで音源分離を行う．

分離信号 $\boldsymbol{y}$ を求めるための分離行列 $\boldsymbol{W}_i$ の更新式は次のようになる．

$$r_{ij,m} = \Sigma_k z_{mk} t_{ik} v_{kj} \tag{8}$$
$$V_{i,m} = \frac{1}{J} \Sigma_j \frac{1}{r_{ij,m}} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^h \tag{9}$$
$$\boldsymbol{w}_{i,m} \leftarrow (\boldsymbol{W}_i V_{i,m})^{-1} \boldsymbol{e}_m \tag{10}$$

ここで，$\boldsymbol{e}_m$ は m 番目の要素のみが 1 の単位ベクトルである．

Rank-1 MNMF の分配関数 z_{mk}，基底行列の要素 t_{ik}，アクティベーション行列の要素 v_{kj} の更新には，前述した分配関数 z_{mk} を使って各分離音源に自動的に基底を割り当てる方法と，$z_{mk} \in \{0, 1\}$ にし，すべての音源を同じ数ずつの基底で表現することで分配関数 z_{mk} を使用しない方法がある．分配関数 z_{mk} を使用しない方法では，z_{mk} の更新がないため，チャネル m ごと

Table 1 実験条件

サンプリング周波数		16 kHz
分析フレーム長		1024, 2048, 4096, 8192 sample
分析フレームシフト幅		フレーム長/4
Rank-1 MNMF の基底数	分配関数なし	1, 5, 10, 15, 20 個
	分配関数あり	8, 40, 80, 120, 160 個
反復更新回数		200 回
入力 SN 比		0, −5, −10 dB

に NMF の更新式を適用することで，$t_{ik,m}$ および $v_{kj,m}$ の更新を行う．その更新式は次式のようになる．

$$t_{il,m} \leftarrow t_{il,m}\sqrt{\frac{\Sigma_j |y_{ij,m}|^2 v_{lj,m}\left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-2}}{\Sigma_j v_{lj,m}\left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-1}}} \quad (11)$$

$$v_{il,m} \leftarrow v_{il,m}\sqrt{\frac{\Sigma_i |y_{ij,m}|^2 t_{il,m}\left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-2}}{\Sigma_i t_{il,m}\left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-1}}} \quad (12)$$

一方，分配関数 z_{mk} を使用して基底を割り当てる方法では，MNMF と同様に z_{mk} の更新も行う必要がある．更新式は次のようになる．

$$z_{mk} \leftarrow z_{mk}\sqrt{\frac{\Sigma_{i,j} |y_{ij,m}|^2 t_{ik} v_{kj}\left(\Sigma_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\Sigma_{i,j} t_{ik} v_{kj}\left(\Sigma_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \quad (13)$$

$$t_{ik} \leftarrow t_{ik}\sqrt{\frac{\Sigma_{j,m} |y_{ij,m}|^2 z_{mk} v_{kj}\left(\Sigma_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\Sigma_{j,m} z_{mk} v_{kj}\left(\Sigma_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \quad (14)$$

$$v_{kj} \leftarrow v_{kj}\sqrt{\frac{\Sigma_{i,m} |y_{ij,m}|^2 z_{mk} t_{ik}\left(\Sigma_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-2}}{\Sigma_{i,m} z_{mk} t_{ik}\left(\Sigma_{k'} z_{mk'} t_{ik'} v_{k'j}\right)^{-1}}} \quad (15)$$

以上より，$\boldsymbol{W}_i$ の更新と，分配関数 z_{mk}，基底行列の要素 t_{ik}，アクティベーション行列の要素 v_{kj} の更新を交互に反復して行うことによって分離信号を求めるための $\boldsymbol{W}_i$ を得る．最後に projection back [7] を適用することで信号のスケールを修正する．

4 評価実験

4.1 実験条件

実際に柔軟索状ロボットを用いて収録した音声のエゴノイズ除去を，定量評価ができるよう再現し，実験を行った．具体的には，災害現場を想定したセットで，8 個のマイクロホンと 7 個の振動モータが取り付けられた全長 3 m の柔軟索状ロボットを用いて，被災者から 8 個のマイクロホンまでのインパルス応答を実測した．このインパルス応答と音声を畳み込んで音源信号とし，SN 比を調整したエゴノイズと加算することによって，シミュレーション用の混合音を作成した．混合音を Rank-1 MNMF で分離し，分離前と分離後を比較し評価した．評価指標には SDR (signal-to-distortion ratio) [8] を用いた．SDR は出力音の歪みの少なさを評価する尺度であり，値が大きいほど音声の分離性能が優れていることを示す．また実際の音源数は未知であるが，音源数とマイクロホン数は同じと仮定した．その他の実験条件を Table 1 に示す．

実験では，まずエゴノイズ除去に用いるのに適切な Rank-1 MNMF の各パラメータについて調査した．そして，SDR 改善量を用いて，柔軟索状ロボットのエゴノイズ除去を行うために Rank-1 MNMF が有効であるかどうかを確認した．

4.2 実験結果と考察

4.2.1 分析フレーム長に関する実験結果と考察

基底数を固定し，分析フレーム長を変えて実験した結果の SDR を図 3 に示す．基底数は，分配関数なしの場合も分配関数ありの場合もそれぞれ最も良い結果が得られたものを示している．分配関数なしの場合では，基底数は各音源に 15 個ずつ割り当てるよう固定されている．一方，分配関数ありの場合では，全ての音源での基底数の総和を 40 個に固定し，分配関数によって各音源に割り当てている．図 3 は音声とエゴノイズの入力 SN 比が −5 dB のときの結果であり，SDR は分析フレーム長 2048 sample のときに最も高い．Rank-1 MNMF の場合，良い結果を導く分析フレーム長は被災者から 8 個のマイクロホンまでのインパルス応答によって変化すると考えられ，この実験のように音声との距離が 1–3 m のときには分析フレーム長を 2048 samples にすることで SDR を大きく改善できることが分かった．

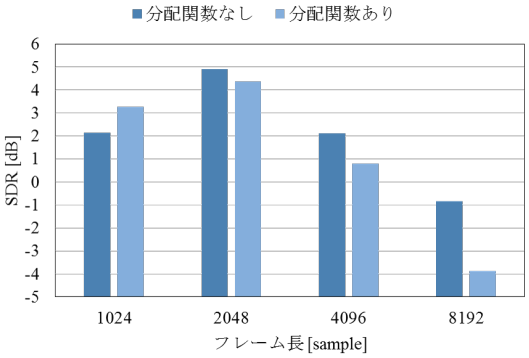

Fig.3 分析フレーム長別の SDR (入力 SN 比: −5 dB)

4.2.2 分配関数による基底の割り当てに関する実験結果と考察

分析フレーム長を 2048 samples に固定し，分配関数の有無および基底の個数を変えて実験した．音声とエゴノイズの入力 SN 比が −5 dB のときの分配関数なしの結果を図 4，分配関数ありの結果を図 5 に示す．また，分配関数ありの場合に 40 個の基底を 8 個の音源に割り当てた分配関数の例を図 6 に示す．図 6 の縦軸は音源インデックス，横軸は基底インデックスであり，分配関数 z_{mk} の値の大きさをそれぞれ色で示している．図 4–5 より，各音源に同じ個数の基底を割り当て，分配関数を使わない場合，SDR は基底数が 15 個のとき最も高いことが分かる．一方，分配関数で割り当てを行う場合，全部で 40 個の基底を割り当てたとき最も SDR の値が高かった．

4.2.3 Rank-1 MNMF の有効性に関する実験結果と考察

入力 SN 比ごとに IVA，Rank-1 MNMF の分配関数あり／なしの場合の SDR 改善量を図 7 に示す．SDR 改善量は，音源分離前の音声に対する SDR を音源分離後の音声に対する SDR から引いたものである．分析フレーム長を 2048 samples に固定し，Rank-1 MNMF に関しては，分配関数なしの場合は音源ごとに基底数を 15 個，分配関数ありの場合は基底数を全体で 40 個にした．図 7 より，Rank-1 MNMF が IVA よりもエゴノイズ除去性能が高いことが分かり，エゴノイズ除去における Rank-1 MNMF の有効性を確認することができた．また，IVA は基底数 1 の場合に相当するため，各音源を表現するために基底数がある程度必要であることが分かる．さらに，分配関数の有無で比較すると，分配関数を使わない場合のほうが SDR が高く，分配関数があまり有効に作用していないことが分かった．分配関数に音声とエゴノイズに合うよう制約を加えることで，エゴノイズ除去においては分配関数の柔軟性を生かせると考えられる．

5 おわりに

本稿では，災害時対応のための柔軟索状ロボットのエゴノイズ除去に，ブラインド音源分離の手法である Rank-1 MNMF を適用し，その有効性を確認するためシミュレーションデータを用いた評価実験を行った．実験では，まず Rank-1 MNMF でのエゴノイズ除去に適切と考えられる分析フレーム長および基底数を調べた．さらに，そのパラメータで Rank-1 MNMF と IVA の結果を比較したところ，いずれの入力 SN 比の場合も Rank-1 MNMF の方が分離精度が高く，エゴノイズ除去への Rank-1 MNMF の適用の有効性を確認した．

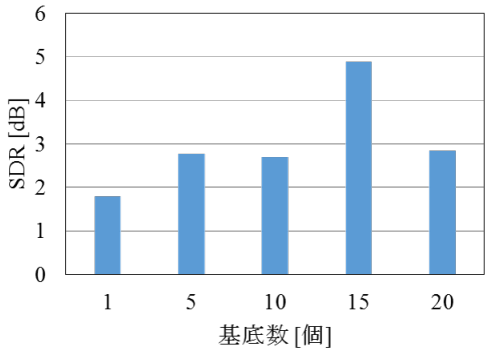

Fig.4 基底数別の SDR (分配関数なし，入力 SN 比: −5 dB)

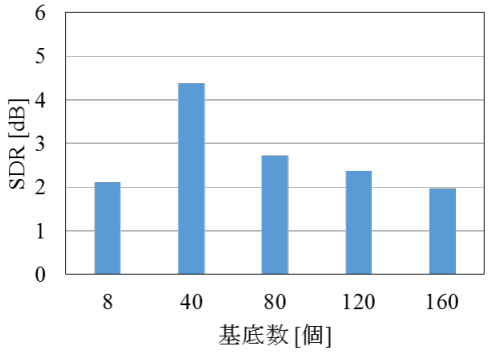

Fig.5 基底数別の SDR (分配関数あり，入力 SN 比: −5 dB)

謝辞

本研究は，総合科学技術・イノベーション会議により制度設計された革新的研究開発推進プログラム (ImPACT) により，科学技術振興機構を通して委託されたものである．実験データを提供して頂いた早稲田大学奥乃博教授と京都大学坂東宜昭氏に感謝の意を表する．

参考文献

[1] ImPACT 革新的研究開発推進プログラム「タフ・ロボティクス・チャレンジ」, http://www.jst.go.jp/impact/program07.html.

[2] H. Namari, K. Wakana, M. Ishikura, M. Konyo and S. Tadokoro, “Tube-type active scope camera with high mobility and practical functionality,” Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 3679–3686, 2012.

[3] D. Kitamura, N. Ono, H. Sawada, H. Kameoka and H. Saruwatari, “Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model,” Proc. ICASSP, pp. 276–280, 2015.

[4] 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.

[5] N. Ono, “Stable and fast update rules for independent vector analysis based on auxiliary function technique,” Proc. WASPAA, pp. 189–192, 2011.

[6] 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.

[7] 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.

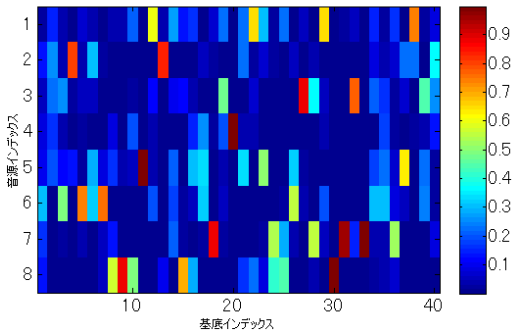

Fig.6 分配関数の例

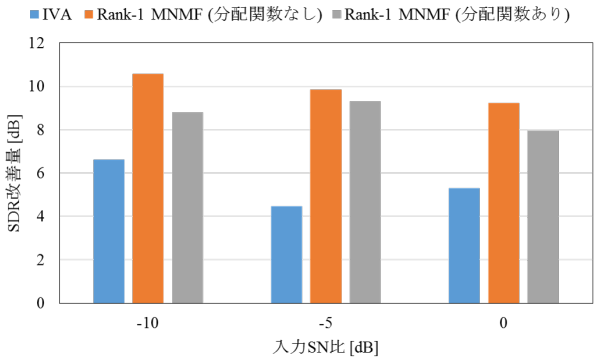

Fig.7 手法別の SDR 改善量

[8] 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.