マルチチャネル非負値行列因子分解における階層的クラスタ分析を用いた初期値設定法*

☆浦本昂伸 (大分大), 太刀岡勇気†, 成田知宏 (三菱電機),
三浦伊織, 上ノ原進吾, 古家賢一 (大分大)

1 はじめに

非負値行列因子分解 (Nonnegative Matrix Factorization: NMF)[1] とは、非負値の行列を分解し、解析を行う手法である。行列表現できるデータならば適用可能なため、音や画像、文書など多種多様なものに利用できる。音響分野では、NMF をマルチチャネル拡張することで空間情報を活用し、音源分離を行うマルチチャネル NMF (MNMF) が提案されている[2, 3]。しかし、MNMF は自由度が高いため、局所最適解に陥りやすく、分離性能の初期値依存性が課題となっている[4]。また、本稿の実験や[7]で示すように、チャネル数が増加するほど、この初期値依存性が顕在化し、音源分離が困難となる。

文献[7]では、混合前信号を教師情報として用いることで、初期値の設定を行った。本稿は、MNMF に有効な初期値設定法として、階層的クラスタ分解を使って得られたクラスタからパラメータを算出し、混合前音源信号を教師情報として利用しない手法を提案する。初期値にランダムな値を設定する従来法に対して、分離性能の比較を行い、提案法の有効性を検証する。

2 MNMF

2.1 概要

MNMF[2, 3] とは、NMF をマルチチャネル拡張したものであり、複素観測行列 $\mathbf{X}$ を 4 つの行列 $\mathbf{H}$,$\mathbf{Z}$,$\mathbf{T}$,$\mathbf{V}$ に分解する。MNMF では空間情報を用いてスペクトル基底を L 個の音源にクラスタリングすることで、事前の学習なしで音源分離を実現する。位相情報を扱うために、複素数における非負性に対応するエルミート半正定値行列を用いる[2]。

2.2 定式化

M をマイクロホン数として入力ベクトルを $\tilde{\mathbf{x}} = [\tilde{x}_1, \cdots, \tilde{x}_M]^\top$ とする。ただし、$^\top$ は転置を表す。$\tilde{x}_m$ は m 番目のマイクロホンでの Short-Time Fourier Transform (STFT) の複素係数であり、スペクトログラムを指す。周波数 i $(1 \leq i \leq I)$、時間 j $(1 \leq j \leq J)$ のとき $\tilde{\mathbf{x}}_{ij}$ で表すと行列 $\mathbf{X}$ の i,j 成分を $\mathsf{X}_{ij} \in \mathbb{C}^{M\times M}$ とし、$\mathsf{X}_{ij} = \tilde{\mathbf{x}}_{ij}\tilde{\mathbf{x}}_{ij}^H$ について

$$\mathsf{X}_{ij} = \tilde{\mathbf{x}}_{ij}\tilde{\mathbf{x}}_{ij}^H = \begin{bmatrix} |\tilde{x}_1|^2 & \cdots & \tilde{x}_1\tilde{x}_M^* \\ \vdots & \ddots & \vdots \\ \tilde{x}_M\tilde{x}_1^* & \cdots & |\tilde{x}_M|^2 \end{bmatrix} \quad (1)$$

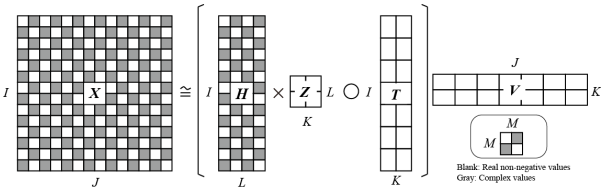

Fig. 1 MNMF で分解された行列の例 (グレーは複素数)

で表される。ただし、^{H} はエルミート転置を表す。すなわち、I 行 J 列の行列 $\mathbf{X}$ は要素が複素行列となる階層的なエルミート半正定値行列である。この行列 $\mathbf{X}$ を MNMF で分解すると、K 個の基底から成る基底行列 $\mathbf{T}(\in \mathbb{R}^{I\times K})$、アクティベーション行列 $\mathbf{V}(\in \mathbb{R}^{K\times J})$、音源の空間情報を示す空間相関行列 $\mathbf{H}$ と音源の空間情報と各基底を関連付ける潜在変数行列 $\mathbf{Z}(\in \mathbb{R}^{L\times K})$ という 4 つの行列の積 $\hat{\mathbf{X}}$ に分解され、次式で示される。

$$\mathbf{X} \approx \hat{\mathbf{X}} = (\mathbf{HZ} \circ \mathbf{T})\mathbf{V} \quad (2)$$

ただし、$\circ$ はアダマール積を表す。行列 $\mathbf{H}$ は行列 $\mathbf{X}$ と同様にそれぞれの要素が $M \times M$ の複素行列を持つ I 行 L 列の階層的なエルミート半正定値行列である。Fig. 1 は式 (2) を図式化したもので、この時、右辺は

$$\hat{\mathsf{X}}_{ij} = \sum_{k=1}^{K}\left(\sum_{l=1}^{L}\mathsf{H}_{il}z_{lk}\right)t_{ik}v_{kj} \quad (3)$$

と表すことができ、理想的には行列 X と $\hat{\mathsf{X}}_{ij}$ を要素に持つ行列 $\hat{\mathbf{X}}$ は等しくなる。しかし、一般的には誤差が生じるため、MNMF では行列 X と行列 $\hat{\mathsf{X}}$ との距離 $D_*(\mathsf{X}, \hat{\mathsf{X}})$ を定義し、この距離を最小化する行列 $\mathbf{T}$, $\mathbf{V}$, $\mathbf{H}$, $\mathbf{Z}$ を求める。今回はダイナミックレンジが大きい音楽や音声に適している Itakura-Saito (IS) divergence[5] を用いて以下のように定義する。

$$D_{IS}(\mathsf{X}_{ij}, \hat{\mathsf{X}}_{ij}) = tr(\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1}) - \log\det\mathsf{X}_{ij}\hat{\mathsf{X}}_{ij}^{-1} - M \quad (4)$$

ただし、$tr\,(\cdot)$ は対角要素の和を表している。

2.3 行列分解アルゴリズム

$D_{IS}(\mathsf{X}, \hat{\mathsf{X}})$ を最小化するために、Multiplicative update rule[6] と呼ばれる反復アルゴリズムを、ランダムな非負の値で初期化した行列 $\mathbf{T}$，$\mathbf{V}$，$\mathbf{Z}$ ならびに各

---

*Initial value setting method using hierarchical cluster analysis for multi-channel non-negative matrix factorization. by Takanobu Uramoto (Oita University), Yuuki Tachioka, Tomohiro Narita (Mitsubishi Electric), Iori Miura, Shingo Uenohara, and Ken'ichi Furuya (Oita University)

†2017 年 4 月三菱電機退職

要素へ単位行列を持たせた行列 $\mathbf{H}$ に繰り返し適用する。IS divergence を用いた場合、更新式は以下のようになる。

$$t_{ik} \leftarrow t_{ik} \sqrt{\frac{\sum_l z_{lk} \sum_j v_{kj} tr(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{il})}{\sum_l z_{lk} \sum_j v_{kj} tr(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{il})}} \quad (5)$$

$$v_{kj} \leftarrow v_{kj} \sqrt{\frac{\sum_l z_{lk} \sum_i t_{ik} tr(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{il})}{\sum_l z_{lk} \sum_i t_{ik} tr(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{il})}} \quad (6)$$

$$z_{lk} \leftarrow z_{lk} \sqrt{\frac{\sum_{i,j} t_{ik} v_{kj} tr(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{il})}{\sum_{i,j} t_{ik} v_{kj} tr(\hat{\mathsf{X}}_{ij}^{-1} \mathsf{H}_{il})}} \quad (7)$$

H_{il} については次式の A，B を係数に持つ代数リッカチ方程式 $H_{il}AHil = B$ を解くことで求めることができる。

$$A = \sum_k z_{lk} t_{ik} \sum_j v_{kj} \hat{\mathsf{X}}_{ij}^{-1} \quad (8)$$

$$B = \mathsf{H}'_{il} \left(\sum_k z_{lk} t_{ik} \sum_j v_{kj} \hat{\mathsf{X}}_{ij}^{-1} \mathsf{X}_{ij} \mathsf{X}_{ij}^{-1} \right) \mathsf{H}'_{il} \quad (9)$$

ただし、H'_{il} は更新前の行列 H_{il} を表しており、解き方は文献 [2] に示されている。

2.4 正規化

行列 $\mathbf{H}$ と行列 $\mathbf{Z}$ については、更新毎に発散を防ぐために正規化を行わなければならない。正規化は以下の式で行った。

$$\mathsf{H}_{il} = \frac{\mathsf{H}_{il}}{tr(\mathsf{H}_{il})}, \quad z_{lk} = \frac{z_{lk}}{\sum_l z_{lk}} \quad (10)$$

2.5 音源分離

式 (11) で表されるウィナーフィルタにより、各音源に対応した分離信号を得られる [2]。

$$\tilde{y}_{ij}^{(l)} = \left(\sum_{k=1}^{K} z_{lk} t_{ik} v_{kj} \right) \mathsf{H}_{il} \hat{\mathsf{X}}_{ij}^{-1} \tilde{\mathbf{x}}_{ij} \quad (11)$$

3 チャネル数増加に伴う初期値依存性

MNMF は自由度の高いモデルであるため、局所最適解が増え、分離性能の初期値依存性が問題となる [4]。さらに、チャネル数を増加させることで、これがより顕著になることが報告されている [7]。ここでは、チャネル数を増加させた場合の初期値依存性について実験的に分析を行う。

3.1 実験条件

実験に用いた混合信号は Table 1[8] の音楽データに、Fig. 2 の環境で測定した RWCP 実環境音声・音響データベースのインパルス応答 (E2A)[9] を畳み込み作成した。Fig. 2 においてマイクロホンは右から順に 1-14 まで番号が付いている。今回の実験で使用したマイクロホン番号を Table 2 に示す。分離処理に用いたパラメータを Table 3 に示す。また、MNMF での IS divergence の計算 (4) において行列式が 0 になるのを防ぐために、X_{ij} の対角要素に 10^{-10} を足している。プログラムは Sawada らのアルゴリズム [2] を MATLAB で実装した。ただし、音源数は既知として pairwize-merge は導入せず、Multiplicative update rule の反復適用のみ行っている。また、文献 [2] に倣い、初めの 20 回は空間相関行列 $\mathbf{H}$ と潜在変数行列 $\mathbf{Z}$ を更新せず、その他の変数のみを更新した。一様分布から生成した生成した 10 個の初期値パターンを用意し、音源分離を実行する。ただし、文献 [2] と同様、$\mathbf{H}$ には各要素の対角成分が $1/M$ の対角行列を持たせ、$\mathbf{Z}$ は 0.2 から 0.4 の一様乱数の値を持たせた。分離性能の評価基準は次式の Signal-to-Distortion Ratio (SDR)[3] を用いた。

$$\mathrm{SDR} = 10 \log_{10} \frac{\sum_t s^{\mathrm{img}}(t)^2}{\sum_t y^{\mathrm{spat}}(t)^2 + y^{\mathrm{int}}(t)^2 + y^{\mathrm{artif}}(t)^2}$$

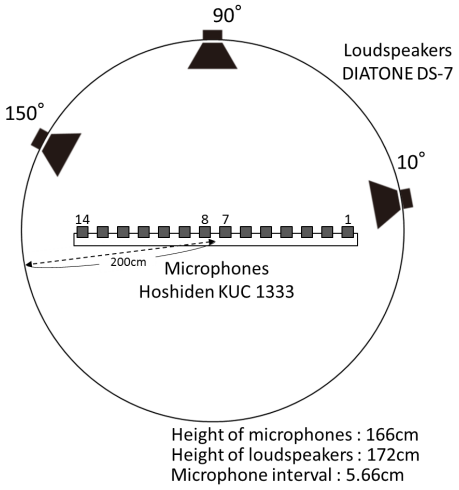

Fig. 2 マイクロホンと音源の配置図

Table 1 実験に用いた音楽データ

ID	Author/Song	Snip	Part
1	Bearlin Roads	85-99 (14 sec)	piano ambient vocals
2	Another Dreamer The Ones We Love	69-94 (25 sec)	drums vocals guitar
3	Fort Minor Remamber The Name	54-78 (24 sec)	drums vocals violin_synth
4	Ultimate Nz Tour	43-61 (18 sec)	drums guitar synth

Table 2 チャネル毎のマイクロホン番号

2ch	6,8
3ch	6,8,10
4ch	4,6,8,10
5ch	4,6,8,10,12
6ch	2,4,6,8,10,12

Table 3 分離処理に用いるパラメータ

残響時間	300ms
サンプリング周波数	16kHz
フレームサイズ	1024
シフトサイズ	256
基底数	30
音源数	3
更新回数	500

ただし、s^{img} は目的音源の正解信号、y^{spat} は空間（フィルタリング）歪み、y^{int} は目的音源以外の音源の信号、y^{artif} は分離処理による信号の歪みを表す。

3.2 初期値依存性 [7]

初期値ランダムの従来法において、単純にマイクロホン数を増やした場合の分離性能を示す。Fig. 3 は各音楽データとチャネル毎の分離後における 3 音源の平均 SDR を示したものである。エラーバーは標準偏差を示す。この図から、3 チャネル以上になっても分離性能が必ずしも向上していないことがわかる。これは、局所最適解による初期値依存性がチャネル数増加に伴って、顕在化するためと考えられる。

4 提案手法

MNMF の分離性能は，空間相関行列 **H** に対する初期値依存性が大きいことが分かっている [4] ため、**H** に着目する。ここでは、初期値がランダムの従来法で得られた分離後の **H** に対して、SDR が最も高いものとその他との関係性について、2 つの **H** の非対角成分の位相差の絶対値を距離 [1] として用いて分析を行った。Fig. 4 では、最も SDR が高いものを best**H** と表記した。best**H** から離れるほど SDR が低下する傾向が見られる。Table 4 に示されるように、各チャネルで距離と SDR には高い負の相関が見られる。そこで、これらの関係性に着目し、分離後の **H** に対して、上記距離に基づいて、階層的クラスタ分析を行う。分析結果から、新たな **H** を計算し、初期値に設定して分離を行う。SDR を計算するには、教師情報として混合前音源信号が必要である。著者らの先行研究 [7] により、教師情報を用いれば分離性能が向上することが予想されるが、事前にそれらの情報を取得することは困難である。そこで、本稿では教師情報を用いない初期値設定法を提案する。

4.1 階層的クラスタ分析

階層的クラスタ分析とは、数値分類法の一種である。異なる集団に属する複数の個体から個体間の距離に基づいて、類似するものを順次集めてクラスタを作成する手法である。クラスタが形成される様子を Fig. 5 のようなデンドログラム [2] で示すことができる。ただ分類するだけではなく、結果として出力されるデンドログラムから任意の数のクラスタに分類することが可能である。例えば Fig. 3 を 3 つのクラスタに分類する場合は、縦線を横に切るような線を引き、その線から下に繋がっている葉を 1 つのクラスタとする。なお、クラスタ間の距離計算にはウォード法 [3] を使用した。

[1] これは距離の公理を満たす。

[2] 木構造に似ているグラフで、ラベルが付いている箇所を葉と言い、葉から伸びている線が連結するまでの高さが短いほど個体が類似している。

[3] 2 つのクラスタを結合した時にクラスタ内の分散が小さく、かつクラスタ間の分散の比を最大化する基準でクラスタを形成する手法。

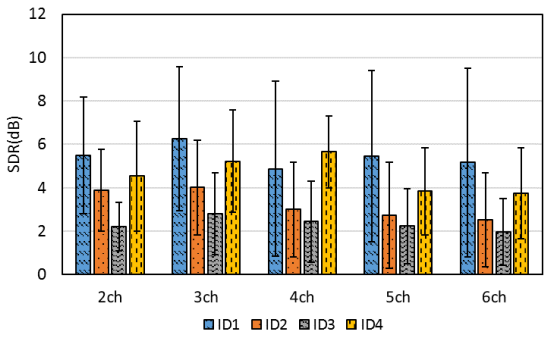

Fig. 3 チャネル数増加に伴う初期値依存性

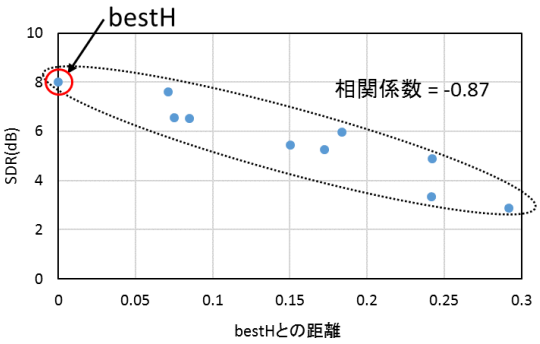

Fig. 4 分離後の行列 H の bestH との距離と SDR の関係性 (2ch)

Table 4 チャネル毎の相関係数

2ch	3ch	4ch	5ch	6ch
-0.87	-0.91	-0.94	-0.91	-0.86

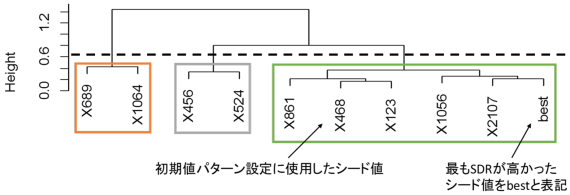

Fig. 5 クラスタ分類の例

4.2 空間相関行列 H の計算

初期値パターンが 10 個であるため、クラスタ数を 2〜9 に設定して分析する。音楽データは Table 1 の ID4 を使用した。Fig. 6,7 では、分類の結果、同じクラスタに属する分離音を平均して SDR を算出した。Fig. 6 は 2 チャネルのデンドログラムを示し、最も SDR が高いクラスタを赤色、最も低いクラスタを青色で囲んでいる。SDR の高いクラスタ内の分離結果の平均により、best よりも高い SDR を得ることができた。また、最も SDR の高いクラスタは best を含んでおり、最も SDR が低いクラスタは、これとは離れた位置に存在している。さらに、SDR が最良のクラスタは、分類されたクラスタの中で最大の要素数を含んでいる傾向が見られた (他のチャネルでも同様)。Fig. 7 に要素数最大のクラスタに属する波形同士を合成して計算した SDR の分析結果を示す。組み合わせの中で得られた最良 SDR を赤色、次点を緑色で表示しており、概ねクラスタ数 2〜4 に分類した時に良い SDR が得られた。そこで、この範囲にクラスタ数を

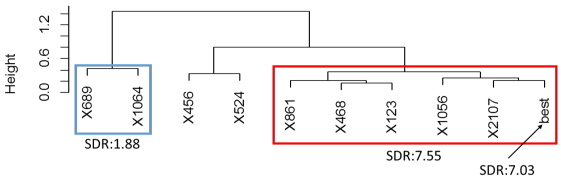

Fig. 6　デンドログラム (2ch)

クラスタ数	チャネル				
	2ch	3ch	4ch	5ch	6ch
2	7.26	8.00	8.57	7.44	7.27
3	7.55	8.20	8.63	7.44	7.34
4	7.55	8.20	8.56	7.35	7.73
5	7.00;7.41	8.10	8.56	7.20	7.73
6	7.00;7.41	7.90	8.18	6.92;4.18	7.50
7	7.00	8.18	8.18	6.92	6.63
8	7.00;7.03	8.57;6.25	8.02	6.67;4.15	6.63
9	7.00	8.57	8.31	6.67	7.11
Best(従来法)	7.03	9.26	8.01	6.58	6.94

Fig. 7　クラスタ数と SDR(dB) の関係性

設定し、要素数最大のクラスタに属する分離後の **H** 同士を平均して新たな **H** を計算する。提案法の流れを Fig. 9 に示す。SDR を計算せずに、クラスタ分析から得られた結果のみを利用して、教師情報を利用しない初期値の設定が可能になる。

5　実験

提案法の有効性を確認するために、従来法と比較実験を行う。実験条件は 3 節と同じである。本実験では、クラスタ数を 3 に固定して空間相関行列 **H** を計算し、MNMF の初期値に設定することで分離を行う。

Fig. 8 は提案法で分離を行った時の結果である。Fig. 3 の従来法よりも SDR が向上し、標準偏差が小さくなることから、分離性能が向上していることが分かる。

従来法では、チャネル数と共に行列の自由度が増加するため、局所最適解に陥りやすくなる。しかし、良いパラメータを推定できている行列を設定することで、局所最適解に陥るのを避け、SDR が向上したと考えられる．ただし、従来法で得られた分離後の **H** をクラスタリングして初期値を得ているため、クラスタ内に分離性能が悪い時の **H** が含まれると、初期値に悪影響を与えることが考えられる。

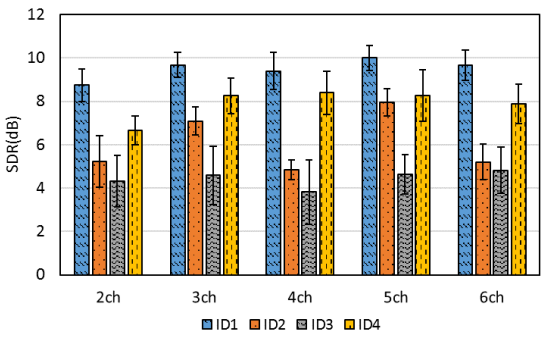

Fig. 8　提案法による実験結果

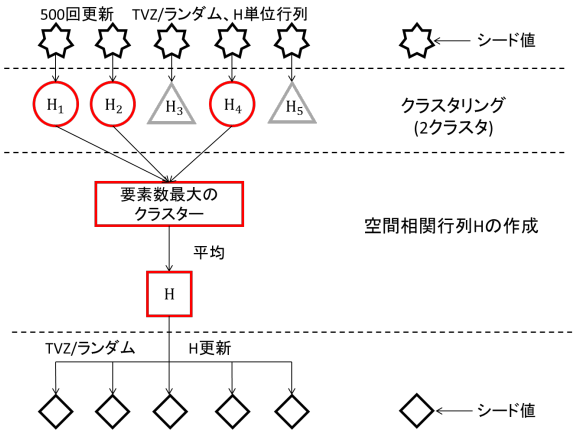

Fig. 9　フローチャート (クラスタ数 2 の例)

6　まとめと今後の課題

本稿では、MNMF のチャネル数増加に伴う初期値依存性を解決するために、階層的クラスタ分析を用いた初期値設定法を提案した。従来法よりも分離性能が向上することから、提案法の有効性を確認した。今後の課題として、以前、提案した逐次的に初期値を設定する手法 [7] との組み合わせが考えられる。

参考文献

[1] D.D. Lee *et al.*, “Learning the Parts of Objects with Nonnegative Matrix Factorization,” Nature, vol.401, pp.788-791, 1999.

[2] H. Sawada *et al.*, “Multichannel Extensions of Non-Negative Matrix Factorization with Complex-Valued Data,” IEEE Trans. ASLP, vol.21, no.5, pp.971-982, 2013.

[3] E. Vincent *et al.*, “First Stereo Audio Source Separation Evaluation Campaigh: Data Algprithm and Results,” Independent Component Analysis and Signal Separation(Springer, Bearlin, 2007), pp.552-559.

[4] 三浦 伊織, 他: “マルチチャネル NMF を用いた音源分離における初期値依存性の挙動解析と音声認識での評価” 信学誌 D, vol.J100-D, pp.376-384, 2017.

[5] C. Fevotte *et al.*, “Nonnegative Matrix Factorization with the Itakura-Saito Divergence: With Application to Music Analysis,” Neural Comput., vol.21, no.3, pp.793-830, 2009.

[6] M. Nakano *et al.*, “Convergence-Guaranteed Multiplicative Algorithms for Non-Negative Matrix Factorization with Beta-Divergence,” In Proc.MLSP 2010, pp.283-288, 2010.

[7] 浦本 昂伸, 他: “マルチチャネル非負値行列因子分解におけるチャネル数増加に伴う逐次的初期値設定法” 音講論集, pp.535-538, 2017.

[8] S. Araki *et al.*, “The 2011 Signal Separation Evaluation Campaign (SiSEC2011): -Audio Source Separation,” Latent Variable Analysis and Signal Separation(Springer, Bearlin, 2012), pp. 414-422.

[9] RWCP: “実環境音声・音響データベース (RWCP-SSD)” 音声資源コンソーシアム, http://research.nii.ac.jp/src/RWCP-SSD.html, 閲覧日:2017/05/31.