Minimal Distortion Principle versus Back Projection for Independent Vector Analysis

Soufiane Tebache, Adel Belouchrani, Lynda Berrah and Nacira Mendjel

Abstract—This paper deals with the scaling ambiguity issue in blind convolutive source separation when performed in the frequency domain. It discusses the relationship between two major techniques, mainly the Minimal Distortion Principle and the Back Projection, that allow to overcome the aforementioned indeterminacy. The Minimal Distortion Principle minimizes the mean square difference between the separated sources and the sensor signals, while the back projection recovers the sensor-observed amplitudes of each estimated source signal. Herein, we prove that the Minimal Distortion Principle is a particular solution of the Back Projection. Another contribution of this paper consists of exploiting one of the most beneficial outcomes of the Back Projection, that is spatial diversity. Our proposed approach applies Single Input Multiple Output deconvolution to the outputs of the back projected source signals, after their estimation by the Independent Vector Analysis algorithm. This method has the advantage of improving the estimation accuracy and removing the channel effect. Experimental results show the effectiveness of our proposal with respect to both the Minimum Distortion Principle and the conventional Back Projection solution.

***Keywords*—Back projection, Minimal Distortion Principle, Independent Vector Analysis, SIMO Deconvolution.**

I. Introduction

Blind Source Separation (BSS) is a powerful tool for separating mixtures of sources from a set of sensors, where neither the source signals nor the mixing parameters are known. This problem leads to two types of indeterminacy: permutation and scaling. When dealing with wide-band propagation, as in the speech signal case, one faces the problem of convolutive mixtures. Existing solutions in the time domain lead to complex and time-consuming computations. Efficient methods solve the underling problem in the time-frequency domain using the Short-Time Fourier Transform (STFT). The latter transforms the convolutive BSS problem into several independent BSS problems with instantaneous models, which can be solved using well-established instantaneous BSS algorithms at each frequency bin. However, the estimated source signals in each frequency bin have arbitrary permutations and scaling that significantly affect the separation performance. Various post-processing [1, 2] can be used to address permutation ambiguity, but this increases computational cost. A more elegant solution is Independent Vector Analysis (IVA) [3], which uses the entire frequency spectrum as input, instead of considering each frequency bin as an independent BSS problem, to overcome the permutation ambiguity.

However, the scaling ambiguity remains and is equivalent to an arbitrary filter of the source signals in the case of the aforementioned approach. The Minimum Distortion Principle (MDP) [9] is a well-known method for dealing with such ambiguity, it selects the separators that minimizes the mean square difference between the separated source signals and the sensor signals. Another technique, called Back Projection (BP) [16], can recover the sources after separation to their sensor-observed amplitudes.

In this paper, we discuss the relationship between Minimal Distortion Principle (MDP) and Back Projection and prove that the MDP is a particular solution of the Back Projection. We observe that the Back Projection actually provides spatial diversity, which has not been exploited yet in the literature. Our second contribution consists of using Single Input Multiple Output (SIMO) deconvolution after separation. This is performed through IVA for each output signal in order to take advantage of the spatial diversity provided by the Back Projection. Such a proposal allows performance enhancement in terms of signal to distortion ratio with respect to reverberate time.

The paper is organized as follows: Section II. formulates the BSS problem briefly, the proposed solution and the proof of the MDP-BP relationship are provided in Section III., simulation results are presented in Section IV., and finally, Section V. concludes the paper.

II. Problem formulation

Let us observe, at time instant n, M signals, $\mathbf{x}(n) = [x_1(n), \cdots, x_M(n)]$, assumed to be the mixtures of L independent source signals $\mathbf{s}(n) = [s_1(n), \cdots, s_L(n)]$ according to the following noiseless convolutive model:

$$x_m(n) = \sum_{l=1}^{L} \sum_{p=0}^{P-1} a_{ml}(p) s_l(n-p) \quad m = 1, .., M \tag{1}$$

$$\mathbf{x}(n) = \sum_{p=0}^{P-1} \mathbf{A}(p)\mathbf{s}(n-p) \tag{2}$$

where $\mathbf{A}$(p), $p = 0, ..., P-1$, is the $M \times L$ transfer function matrix, whose elements are denoted $a_{ml}(p)$, and P is the impulse

Manuscript received October 8, 2025; revised November 7, 2025.
S. Tebache, A. Belouchrani, L. Berrah and N. Mendjel are with Ecole Nationale Polytechnique/LDCCP lab., Algiers, ALGERIA. (e-mail: soufiane.tebache@g.enp.edu.dz, adel.belouchrani@g.enp.edu.dz, lynda.berrah@g.enp.edu.dz, and nacira.mendjel@g.enp.edu.dz)
Digital Object Identifier (DOI): 10.53907/enpesj.v5i2.342

Part of this paper was presented in the conference paper [18].

response length. By assuming a significantly longer time analyzing window than the impulse response length, it is possible to express the convolution in the frequency domain as individual multiplications for each frequency bin, as follows:

$$\mathbf{x}_{TF}(f,n) = \mathbf{A}(f)\mathbf{s}_{TF}(f,n) \tag{3}$$

where $\mathbf{x}_{TF}(f,n)$ represents a column vector with M elements, denoted $x_{TF_m}(f,n)$, which corresponds to the $(f,n)^{th}$ element of the Short-Time Fourier Transform (STFT) [14], $\mathbf{X}_{TF_m}$, of the sensor signal $x_m(t)$, $\mathbf{s}_{TF}(f,n)$ denotes a column vector with L elements, denoted $s_{TF_l}(f,n)$, which corresponds to the $(f,n)^{th}$ element of the STFT, $\mathbf{S}_{TF_l}$, of the source signal $s_l(t)$. The matrix $\mathbf{A}(f)$ is an $M \times L$ mixing matrix for frequency bin f. In order to separate the source signals from the observed mixtures, an unmixing matrix $\mathbf{W}(f)$ should be estimated for each frequency bin:

$$\mathbf{y}_{TF}(f,n) = \mathbf{W}(f)\mathbf{x}_{TF}(f,n) \tag{4}$$

In the sequel, we assume that $M = L$ and that the $\mathbf{A}(f)$ matrices are well conditioned.

III. Proposed Approach

The proposed solution starts by applying the Independent Vector Analysis (IVA) [3] that solves implicitly the permutation ambiguity at the frequency bins, then addresses the scaling ambiguity through Back Projection and finally, recovers the source signal through a SIMO deconvolution algorithm. In subsection B., we discuss the relationship between Minimal Distortion Principle (MDP) and Back Projection, and show that the MDP is a particular solution of the Back Projection.

A. Independent Vector Analysis (IVA)

The IVA algorithm [3] considers the sources as multidimensional random vectors containing all the frequency components of each source signal. It aims to maximize the independence between individual source signals while maintaining the dependency within each vector. This process get rid of the permutation ambiguity between the frequency bins. The unmixing matrix is estimated at each frequency bin according to the following iterative update [11]:

$$\mathbf{W}(f) \leftarrow \mathbf{W}(f) + \eta\Delta\mathbf{W}(f) \tag{5}$$

With

$$\begin{aligned}
\Delta\mathbf{W}(f) &= \left\{\mathbf{I} + E\left[\boldsymbol{\varphi}^f(\mathbf{y}_{TF})\mathbf{y}_{TF}(f)^H\right]\right\}\mathbf{W}(f)\\
\mathbf{y}_{TF} &= [\mathbf{y}_{TF_1},\ldots,\mathbf{y}_{TF_L}]^T\\
\boldsymbol{\varphi}^f(\mathbf{y}_{TF}) &= \left[\varphi_1^f(\mathbf{y}_{TF_1}),\ldots,\varphi_L^f(\mathbf{y}_{TF_L})\right]^T\\
\varphi_l^f(\mathbf{y}_{TF_l}) &= \frac{\partial}{\partial y_{TF_l}(f)}\log\left(p(\mathbf{y}_{TF_l})\right)
\end{aligned}$$

where the step size, $\eta \in [0,1]$, is a tuning parameter that imposes a trade-off between convergence speed and stability [11], and $p(.)$ is a density probability function.

This can be obtained by dimension reduction at each frequency bin, e.g. by data whitening.

B. Minimal Distortion Principle versus Back Projection

To deal with the scaling ambiguity in the frequency domain, two major techniques exist in the literature, mainly the Minimal Distortion Principle (MDP) [9] and the Back Projection (BP) [16]. The MDP chooses the proper separators that minimize the mean square difference between the separated sources and the sensor signals, its solution consists of the following unmixing matrix at frequency bin f:

$$\mathbf{W}_s(f) = diag(\mathbf{W}^{-1}(f))\mathbf{W}(f) \tag{6}$$

where $diag(.)$ denotes the diagonal matrix operator and $\mathbf{W}(f)$ is the unmixing matrix, at frequency bin f, estimated by the IVA algorithm.

The Back Projection technique projects back the estimated source signals to the sensor array, it actually rescales the source signals to match their observed amplitudes at each sensor:

Let $\hat{\boldsymbol{A}}(f) = \mathbf{W}^{-1}(f)$ be the estimated mixing matrix, the inverse of the estimated unmixing matrix, at the f^{th} frequency bin, and $y_{TF_l}(f,n)$, be the estimate of the l^{th} source signal $s_{TF_l}(f,n)$ at the time-frequency point (f,n). Because of the inherent scaling indeterminacy, we have the following relationships:

$$y_{TF_l}(f,n) = \alpha_l(f)s_{TF_l}(f,n) \tag{7}$$

$$\hat{\boldsymbol{a}}_l(f) = \frac{1}{\alpha_l(f)}\boldsymbol{a}_l(f) \tag{8}$$

where $\alpha_l(f)$ is an arbitrary factor. To get the signals, denoted $\underline{\mathbf{s}}_{TF_l}(f,n)$, free from the scaling indeterminacy, one can multiply the l^{th} separated signal $y_{TF_l}(f,n)$ by its corresponding column of the estimated mixing matrix $\hat{\boldsymbol{a}}_l(f)$:

$$\underline{\mathbf{s}}_{TF_l}(f,n) = \hat{\boldsymbol{a}}_l(f)y_{TF_l}(f,n) \tag{9}$$

By substituting equations (7) and (8) in equation (9), one gets:

$$\underline{\mathbf{s}}_{TF_l}(f,n) = \boldsymbol{a}_l(f)s_{TF_l}(f,n), l = 1,\cdots,L \tag{10}$$

where $\underline{\mathbf{s}}_{TF_l}(f,n) = [\underline{s}_{TF_{l_1}}(f,n),\ldots,\underline{s}_{TF_{l_M}}(f,n)]^T$ with $\underline{s}_{TF_{l_k}}(f,n)$ being the contribution of the l^{th} estimated source signal at the k^{th} sensor. It appears clearly from equation (10) that the recovered source signals at their sensor-observed amplitudes are free from the scaling ambiguity $\alpha_l(f)$. Note that equation (10) describes a Single Input Multiple Output (SIMO) system for each source, separately. Thanks to the multidimensionality of its output, such a system provides the spatial diversity needed to recover its input (in our case the corresponding source) using only second order statistics.

Let $\mathbf{W}(f)$ be an estimate of the unmixing matrix, at frequency bin f, of the blind convolutive separation problem. The scale indeterminacy free unmixing matrix of the Minimum Distortion Principle,

$$\mathbf{W}_s(f) = diag(\mathbf{W}^{-1}(f))\mathbf{W}(f)$$

is a particular solution of the Back Projection.

Using equation (9), the expression of the l^{th} signal recovered by Back Projection at the l^{th} sensor is given by:

$$\underline{s}_{TF_{l_l}}(f,n) = \hat{a}_{ll}(f)y_{TF_l}(f,n), l = 1,\cdots,L. \tag{11}$$

Consider the following source vector:

$$\underline{\mathbf{s}}_{TF}(f,n) = [\underline{s}_{TF_{1_1}}(f,n), \underline{s}_{TF_{2_2}}(f,n), ..., \underline{s}_{TF_{L_L}}(f,n)]^T \tag{12}$$

Using equation (11), the source vector $\underline{\mathbf{s}}_{TF}(f,n)$ reads:

$$\underline{\mathbf{s}}_{TF}(f,n) = \begin{bmatrix} \hat{a}_{11}(f) & 0 & \dots & 0 \\ 0 & \hat{a}_{22}(f) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{a}_{LL}(f) \end{bmatrix} \mathbf{y}_{TF}(f,n) \tag{13}$$

Note that

$$\begin{bmatrix} \hat{a}_{11}(f) & 0 & \dots & 0 \\ 0 & \hat{a}_{22}(f) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{a}_{LL}(f) \end{bmatrix} = diag(\hat{\boldsymbol{A}}(f)) \tag{14}$$

where $\hat{\boldsymbol{A}}(f)$ is an estimate of the mixing matrix at frequency bin f, that is the inverse of the estimated unmixing matrix $\mathbf{W}(f)$:

$$\hat{\boldsymbol{A}}(f) = \mathbf{W}^{-1}(f) \tag{15}$$

According to equations (4), (14) and (15), equation (13) is rewritten as:

$$\begin{aligned} \underline{\mathbf{s}}_{TF}(f,n) &= diag(\mathbf{W}^{-1}(f))\mathbf{W}(f)\mathbf{x}_{TF}(f,n) \\ &= \mathbf{W}_s(f)\mathbf{x}_{TF}(f,n) \end{aligned} \tag{16}$$

with $\mathbf{W}_s(f) = diag(\mathbf{W}^{-1}(f))(\mathbf{W})(f)$, where one recognizes the MDP unmixing matrix (6).

This Theorem shows that the Minimum Distortion Principle is a particular solution of the Back Projection, in which only the $l - l^{th}$ components, $\underline{s}_{TF_{l_l}}(f,n), l = 1, \cdots, L$ are used.

C. *Back projection-Spatial diversity*

In reference [17], the authors estimate the source signals as the back projected components onto the first sensor, i.e.

$$\underline{s}_{TF_{l_1}}(f,n), l = 1, \cdots, L \tag{17}$$

The above expression does not exploit the spatial diversity offred by the source vectors $\underline{\mathbf{s}}_{TF_l}(f,n), l = 1, \cdots, L$. Our contribution consists of exploiting this spatial diversity through Single Input Multiple Output (SIMO) deconvolution of the back projected output signal obtained after separation in the time-frequency domain by the IVA algorithm.

The source vectors of equation (9) obtained after the back projection of the IVA algorithm outputs in the time-frequency domain are transformed to the time domain through the Inverse Short Time Fourier Transform (ISTFT):

$$\underline{s}_{l_k}(n) = ISTFT(\underline{s}_{TF_{l_k}}(f,n)), k, l = 1, \cdots, L \tag{18}$$

According to equation (10), one has:

$$\underline{s}_{l_k}(n) = ISTFT(a_{l_k}(f) s_{TF_l}(f,n)), \tag{19}$$

Since we have assumed a significantly longer analyzing window than the impulse response, the multiplications in the frequency domain are translated to linear convolutions in the time domain. Hence, we obtain L SIMO systems:

$$\underline{s}_{l_k}(n) = \sum_{p=0}^{P-1} a_{l,k}(p) s_l(n-p), k, l = 1, \cdots, L \tag{20}$$

The Robust Normalized Multichannel Frequency-Domain Least-Mean-Square algorithm [15] and the SIMO equalizer reported in [8] are employed in this paper for performing the blind identification of the L SIMO channels $a_{l,k}(p), p = 0, \cdots, P-1; k, l = 1, \cdots, L$, and their equalization to retrieve the L original sources $s_l(n), l = 1, \cdots, L$, respectively.

IV. Numerical Experiments

Herein, an evaluation is conducted in the case of speech signals.

A. *Experimental setup*

Pyroomacoustics software package [19] is used to generate the Room Impulse Responses (RIRs) and the corresponding convolutive mixtures according to a simulation scenario: A room measuring $5.5\ m \times 3.5\ m \times 3\ m$ with RT60 reverberation time of $130\ ms$ was chosen, and an array of seven microphones was placed in the center of the room, with one microphone in the center and the other six spaced equally around a circle of a $4.5\ cm$ radius. Two sources were positioned at different angles $0.5\ m$ away from the microphone array, and mixtures were produced using $10\ s$ speaking utterances at a sampling frequency of $16\ kHz$. Figure 1 shows the simulation scenario and the position of the two sources.

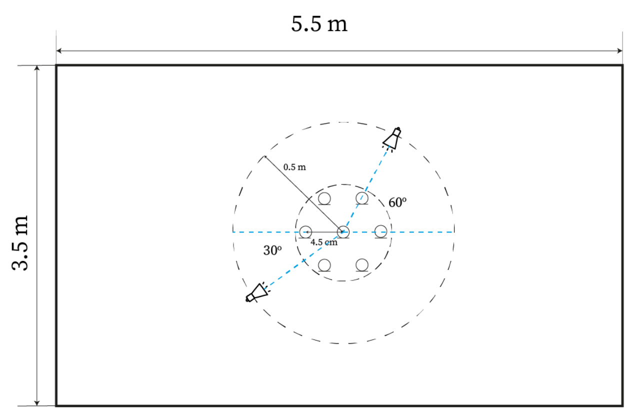

Fig. 1: Simulation scenario with 2 sources.

B. *Performance evaluation*

To assess the quality of the separation, we compute the standard energy ratios in decibels (dB), specifically the Signal-to-Distortion (SDR) for the l^{th} source as:

$$SDR_l = 10 \log_{10} \frac{||s_l||^2}{||s_l - \hat{s}_l||^2} \tag{21}$$

where $||.||$ denotes the Euclidean norm.

RT60 reverberation time is the duration required for the sound energy in a room to decrease by 60 dB after the source emission has stopped (ISO 3382).

Herein, one evaluates the robustness of the proposed IVA-based algorithm versus RT60 reverberation time using MDP [9], BP [17] and our proposal BP-SIMO, respectively. The SDRs of the two sources are computed for different RT60s ranging from 150 ms to 450 ms. Figures 2 and 3 depict the evolution of the BSS performance (SDR) of the first and second separated signal, respectively, as the RT60 increases.

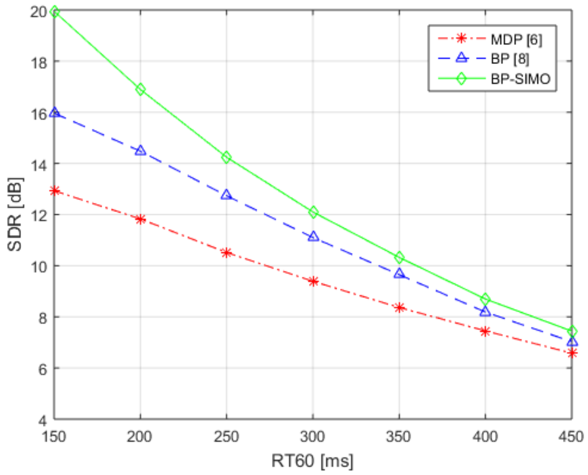

Fig. 2: Effect of reverberation on the SDR of the first separated signal.

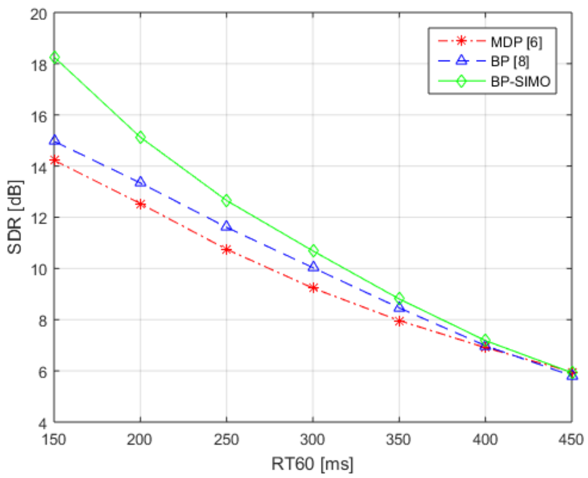

Fig. 3: Effect of reverberation on the SDR of the second separated signal.

The above graphs show that an increase in RT60 reverberate time leads to a gradual degradation of performance. This degradation occurs because sound waves in a reverberate room bounce off surfaces and create multiple reflections that can overlap with the direct sound, making it difficult for BSS algorithms to accurately distinguish between individual sources. The results highlight, as well, that our BP-SIMO approach outperforms the BP [17] and MDP [9] methods for all RT60 values.

V. Conclusion

New insights on the scaling ambiguity problem involved in Independent Vector Analysis is given. The relationship between the Minimal Distortion Principle and Back Projection for solving the aforementioned ambiguity is discussed. The paper shows that the Minimal Distortion Principle is actually a particular solution of the Back Projection. A second contribution consisted of exploiting, through SIMO deconvolution, the spatial diversity provided by the Back Projection. Herein, the SIMO deconvolution has been performed using the Robust Normalized Multichannel Frequency-Domain Least-Mean-Square algorithm [15] for the channel blind identification and the SIMO equalizer reported in [8] for the channel equalization. Performance results, in terms of signal to distortion ratio, confirm that the proposed approach enhances the quality of source separation, with respect to reverberate time, as applied to speech signals.

References

[1] Leandro E. and Diego H. Milone, "Using multiple frequency bins for stabilization of FD-ICA algorithms, Signal Processing, vol. 119, pp. 162-168, 2016.

[2] Yatabe, K., "Consistent ICA : Determined BSS Meets Spectrogram Consistency", IEEE Signal Processing Letters, vol. 27, 2020.

[3] Kim, T. and Lee, I. and Lee, T.-W., "Independent vector analysis: definition and algorithms", 2006 Fortieth Asilomar Conference on Signals, Systems and Computers, pp. 1393–1396, 2006.

[4] Herault, J. and Jutten, C., "Space or Time Adaptive Signal Processing by Neural Network Models, American Institute of Physics Inc., pp. 206–211, 1986

[5] Comon, P., "Independent component analysis, A new concept?", Signal Processing, vol. 36, no. 3, pp. 287-314, 1994.

[6] Comon, P., "Workshop on Higher-Order Spectral Analysis", Separation Of Stochastic Processes, 1989.

[7] Even, J. and Saruwatari, H. and Shikano, K., "Enhanced wiener post-processing based on partial projection back of the blind signal separation noise estimate", 17th European Signal Processing Conference, 2009.

[8] Miyoshi, M. and Kaneda, Y., "Inverse filtering of room acoustics", IEEE Transactions on acoustics, speech, and signal processing, vol. 36, no. 2, pp. 145-152, 1988.

[9] Matsuoka, K., "Minimal distortion principle for blind source separation", Proceedings of the 41st SICE Annual Conference. SICE 2002., vol. 4, pp. 2138–2143, 2002.

[10] Scheibler, R., "Generalized Minimal Distortion Principle for Blind Source Separation", in Proc. Interspeech, Shanghai, China, 2020

[11] Kim, T. and Attias, H.-T. and Lee, S.-Y. and Lee, T.-W., "Blind source separation exploiting higher-order frequency dependencies", IEEE transactions on audio, speech, and language processing, vol. 15, no. 1, pp. 70-79, 2007.

[12] Huang, Yiteng and Benesty, Jacob, "Adaptive blind channel identification: multi-channel least mean square and Newton algorithms", Signal Processing, vol. 82, pp. 1127-1138, 2002.

[13] Xu, Guanghan and Liu, Hui and Tong, Lang and Kailath, Thomas, "A least-squares approach to blind channel identification", IEEE Transactions on signal processing, vol. 43, no. 12, pp. 2982-2993, 1995.

[14] Oppenheim, A. V., "Speech spectrograms using the fast Fourier transform, IEEE Spectrum, vol. 7, no. 8, pp. 57-62, 1970.

[15] Haque, M. A. and Hasan, M. K., "Noise robust multichannel frequency-domain LMS algorithms for blind channel identification", IEEE Signal Processing Letters, vol. 15, pp. 305-308, 2008.

[16] Murata, N. and Ikeda, S. and Ziehe, A., "An approach to blind source separation based on temporal structure of speech signals", Neurocomputing, vol. 41, pp. 1-24, 2001.

[17] Nakashima, T. and Ono, N., "Inverse-free Online Independent Vector Analysis with Flexible Iterative Source Steering", 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2022.

[18] Belouchrani, A. and Mendjel, N. and Berrah, L. and Tebache, S., "Independent Vector Analysis Based MIMO Deconvolution: Exploiting Spatial Diversity Through Back Projection", 22nd IEEE Statistical Signal Processing Workshop, Hanoi, Vietnam, 2-5 July 2023.

[19] Scheibler, R. and Bezzam, E. and Dokmanić, I., "Pyroomacoustics: A python package for audio room simulation and array processing algorithms2, 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP), 2018.

[20] Emmanuel Vincent, "BSS Eval Toolbox Version 3.0 for Matlab", 2007.

[21] Hiroe, Atsuo, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions", 6th International Conference, ICA 2006, Charleston, SC, USA, March 5-8, 2006.

[22] Liu, A. and Song, G. and Lee, S. and Fu, X. and Chen, X., A State-dependent IVA Model for Muscle Artifacts Removal from EEG Recordings, IEEE Transactions on Instrumentation and Measurement, vol. 70, 2021.

[23] Belouchrani, A. and Cichock, A., "Robust whitening procedure in blind source separation context2, IEE Electronics Letters, vol. 36, no. 24, 2000.