Evaluation of Multichannel Hearing Aid System by Rank-Constrained Spatial Covariance Matrix Estimation

Masakazu Une*, Yuki Kubo†, Norihiro Takamune†, Daichi Kitamura‡, Hiroshi Saruwatari† and Shoji Makino*
*University of Tsukuba, Graduate School of Systems and Information Engineering, Ibaraki, Japan
†The University of Tokyo, Graduate School of Information Science and Technology, Tokyo, Japan
‡National Institute of Technology, Kagawa College, Kagawa, Japan

***Abstract*—In a noisy environment, speech extraction techniques make hearing aid systems more effective and practical. Blind source separation (BSS) is suitable for hearing aids because it can be employed without any a priori spatial information. Among many BSS methods, independent low-rank matrix analysis (ILRMA) achieves high-quality separation performance. In a diffuse-noise environment, however, ILRMA cannot suppress the noise since it is based on the determined situation. On the other hand, rank-constrained spacial covariance matrix (SCM) estimation overcomes the problem. This method utilizes spatial parameters accurately estimated by ILRMA and compensates for the deficiency of the spatial basis of diffuse noise. The application of BSS methods to a multichannel binaural hearing aid system with a smartphone has never been studied in detail thus far. To clarify the efficacy of the BSS methods in real environments, we record real sounds by constructing a hearing aid system with a dummy head and a smartphone. In this study, we investigate the applicability of BSS for a multichannel binaural hearing aid system with microphones on a smartphone. Furthermore, we apply ILRMA and the rank-constrained SCM estimation to the recorded data and evaluate these methods in terms of their separation performance.**

I. Introduction

When we use a binaural hearing aid in a noisy environment, target-speech extraction is necessary since speech is always contaminated by noise. In binaural hearing aid systems, blind source separation (BSS) [1] is suitable because it works well without spatial information, e.g., microphone positions around both ears, the head-related acoustic condition, the target-speaker location (direction), and room reverberation. Many BSS methods, such as frequency domain independent component analysis [2], [3], independent vector analysis [4]–[6], and independent low-rank matrix analysis (ILRMA) [7]–[9], have been proposed . In particular, ILRMA achieves effective and accurate separation by introducing nonnegative matrix factorization (NMF) [10] to the source model. However, these methods can be applied to only the determined or overdetermined situations (the number of microphones $\geq$ the number of sources), and their applicability is not realized when these conditions are not satisfied (i.e., underdetermined case). On the other hand, rank-constrained spacial covariance matrix (SCM) estimation [11] has been proposed as an effective method for a situation in which noise arrives from all directions, i.e., diffuse-noise case. Basically, this method estimates a full-rank SCM [12], which represents spatial characteristics of the diffuse noise, just as multichannel NMF (MNMF) [13], [14] does. However, MNMF requires the estimation of an enormous number of parameters, leading to a high computational cost. In contrast, the rank-constrained SCM estimation reduces the number of parameters by using the highly accurate spatial parameters obtained by ILRMA and restores the lost spatial basis for diffuse noise. It was reported that the rank-constrained SCM estimation achieves a more efficient and stable extraction of target speech than MNMF under the simplified computer-simulation-based acoustic condition, but the evaluation in real situations remains a problem.

In this paper, first, we propose a new multichannel hearing aid system composed of a *distributed microphone array* including binaural ear-attached microphones and smartphone microphones. We utilize eight microphones that are synchronized with the same sampling rate. Although all the microphone positions are not specified in advance owing to variations in user head sizes and smartphone location, BSS is fully applicable to the distributed configuration. To the best of our knowledge, there has been no study on BSS applied to such a system. Second, we implement the rank-constrained SCM estimation as a speech extraction algorithm in this hearing aid system. We prepare a head-and-torso dummy with a smartphone to imitate a person (hearing aid user) holding a smartphone, and record the real diffuse noise at the multiple microphones. On the basis of the results, we evaluate the speech extraction performance of the rank-constrained SCM estimation. The experimental results show that the proposed method outperforms the conventional ILRMA in a realistic situation.

II. Formulation and BSS Algorithms

A. Formulation

Let us consider separating the observed signals, which are obtained by M microphones capturing the signals arriving from N sources. The source, observed, and separated signals in each time-frequency slot are denoted as $\boldsymbol{s}_{ij} = (s_{ij,1}, \ldots, s_{ij,N})^\top \in \mathbb{C}^N$, $\boldsymbol{x}_{ij} = (x_{ij,1}, \ldots, x_{ij,M})^\top \in \mathbb{C}^M$, and $\boldsymbol{y}_{ij} = (y_{ij,1}, \ldots, y_{ij,N})^\top \in \mathbb{C}^N$, where $i = 1, \ldots, I$, $j = 1, \ldots, J$, and $n = 1, \ldots, N$ indicate the indexes of the frequency bins, time frames, and sources, respectively.

The operator $\cdot^{\top}$ indicates transpose. When each source is a directional target source and the window length of short-time Fourier transform (STFT) is sufficiently larger than that of the impulse response of the object space, the observed signal and the mixing matrix $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \cdots \boldsymbol{a}_{i,N}) \in \mathbb{C}^{M \times N}$ in each frequency bin have the relation

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}, \tag{1}$$

where $\boldsymbol{a}_{i,n}$ is the steering vector for each source. If the number of microphones is equal to that of sources ($M = N$) and $\boldsymbol{A}_i$ is not a singular matrix, the separated signal $\boldsymbol{y}_{ij}$ can be obtained by estimating the demixing matrix $\boldsymbol{W}_i = \boldsymbol{A}_i^{-1} = (\boldsymbol{w}_{i,1} \cdots \boldsymbol{w}_{i,N})^{\mathrm{H}} \in \mathbb{C}^{N \times M}$ as

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}, \tag{2}$$

where the operator $\cdot^{\mathrm{H}}$ denotes the Hermitian transpose.

B. ILRMA [7]

In ILRMA, the component of the nth source in each time-frequency slot is assumed to be generated from a statistical model that follows the univariate complex Gaussian distribution as

$$s_{ij,n} \sim \mathcal{N}_c \left(0, \sum_l t_{il,n} v_{lj,n} \right), \tag{3}$$

where $t_{il,n} \geq 0$ and $v_{lj,n} \geq 0$ are NMF variables, $l = 1, \ldots, L$ is an index of the NMF basis, and L is the number of bases. Simultaneously, the observed signal $\boldsymbol{x}_{ij}$ follows the multivariate complex Gaussian distribution because of the reproductive property, i.e.,

$$\boldsymbol{x}_{ij} \sim \mathcal{N}_c \left(\boldsymbol{0}, \sum_n r_{ij,n} \boldsymbol{a}_{i,n} \boldsymbol{a}_{i,n}^{\mathrm{H}} \right), \tag{4}$$

$$r_{ij,n} = \sum_l t_{il,n} v_{lj,n}, \tag{5}$$

where $r_{ij,n}$ corresponds to the nth source model that approximates the power spectrogram using nonnegative values $t_{il,n}$ and $v_{lj,n}$. The steering vector $\boldsymbol{a}_{i,n}$ corresponds to the spatial model as the rank-1 SCM constructed by the spatial basis for the nth source. These NMF variables $t_{il,n}$, $v_{lj,n}$ and the demixing matrix $\boldsymbol{W}_i$ are obtained by maximum-likelihood estimation based on the maximization of statistical independence between the sources.

C. Rank-Constrained SCM Estimation [11]

The rank-constrained SCM estimation focuses on a situation where one directional target source and diffuse noise are mixed. As the strategy of this method, the full-rank SCM is estimated using the spatial basis of the one directional target source and the noise SCM is estimated by ILRMA. First, we apply ILRMA to $\boldsymbol{x}_{ij}$ and obtain one a "noise-contaminated target speech" component and $M - 1$ "noise-only" components (see [15] for the physical mechanism of this phenomenon). Second, we calculate the noise SCM using the above-mentioned components. Since the ILRMA-estimated noise SCM lacks a rank corresponding to the target source direction, which results in the rank-($M - 1$) SCM, rank-constrained SCM estimation is used to estimate the parameters to compensate for the lack of the rank. Finally, multichannel Wiener filtering is applied to suppress the noise diffusing toward the target source. The overview of the algorithm is described below.

The rank-constrained SCM estimation assumes the observed signal $\boldsymbol{x}_{ij}$ as the sum of the source image vector $\boldsymbol{h}_{ij} = (h_{ij,1}, \ldots, h_{ij,M})^{\top}$ and the diffuse noise image vector $\boldsymbol{u}_{ij} = (u_{ij,1}, \ldots, u_{ij,M})^{\top}$; i.e.,

$$\boldsymbol{x}_{ij} = \boldsymbol{h}_{ij} + \boldsymbol{u}_{ij}. \tag{6}$$

The source image vector $\boldsymbol{h}_{ij}$ is expressed using a vector corresponding to the target source, $\boldsymbol{a}_i^{(h)} =: \boldsymbol{a}_{i,n_h}$ out of the spatial bases $\boldsymbol{a}_{i,1}, \ldots, \boldsymbol{a}_{i,N}$ obtained by ILRMA, and the target source image $s_{ij}^{(h)}$ as follows:

$$\boldsymbol{h}_{ij} = \boldsymbol{a}_i^{(h)} s_{ij}^{(h)}, \tag{7}$$

$$s_{ij}^{(h)} \sim \mathcal{N}_c \left(0, r_{ij}^{(h)} \right), \tag{8}$$

where n_h indicates the index corresponding to the target source and $r_{ij}^{(h)}$ is the variance of the target source (power spectrogram).

The variance of the target source $r_{ij}^{(h)}$ is assumed to have sparsity and the prior distribution follows an inverse gamma distribution given by

$$p(r_{ij}^{(h)}; \alpha, \beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} \left(r_{ij}^{(h)} \right)^{-\alpha - 1} \exp \left(-\frac{\beta}{r_{ij}^{(h)}} \right), \tag{9}$$

where $\alpha > 0$, $\beta > 0$, and $\Gamma(\cdot)$ are the shape parameter, scale parameter, and gamma function, respectively. On thie other hand, the diffuse noise $\boldsymbol{u}_{ij}$ follows the following multivariate complex Gaussian distribution and is statistically independent of the target source $\boldsymbol{h}_{ij}$:

$$\boldsymbol{u}_{ij} \sim \mathcal{N}_c \left(\boldsymbol{0}, r_{ij}^{(u)} \mathbf{R}_i^{(u)} \right), \tag{10}$$

where $r_{ij}^{(u)}$ and $\mathbf{R}_i^{(u)}$ are the variance and the full-rank SCM of the diffuse noise, respectively. Here, N separated signals $\hat{\boldsymbol{y}}_{ij,1}, \ldots, \hat{\boldsymbol{y}}_{ij,N}$ are obtained by ILRMA, and the SCM of the diffuse noise $\mathbf{R}_i^{(u)}$ is represented as

$$\mathbf{R}_i^{(u)} = \mathbf{R}'^{(u)}_i + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{H}}, \tag{11}$$

$$\mathbf{R}'^{(u)}_i = \frac{1}{J} \sum_j \hat{\boldsymbol{y}}_{ij}^{(u)} \left(\hat{\boldsymbol{y}}_{ij}^{(u)} \right)^{\mathrm{H}}, \tag{12}$$

$$\hat{\boldsymbol{y}}_{ij}^{(u)} = \sum_{n \neq n_h} \hat{\boldsymbol{y}}_{ij,n}, \tag{13}$$

where $\mathbf{R}'^{(u)}_i$ is the noise SCM estimated by ILRMA whose rank is $M - 1$, $\boldsymbol{b}_i$ is a unit eigenvector corresponding to zero eigenvalue of $\mathbf{R}'^{(u)}_i$ and λ_i is the weight variable. Note that, in $\mathbf{R}_i^{(u)}$, only λ_i is the variable to be optimized because $\mathbf{R}'^{(u)}_i$ and $\boldsymbol{b}_i$ are given by ILRMA as fixed values in advance. By

modeling the prior distribution of the target source variance in (9), we express the negative log likelihood function $\mathcal{L}$ of the rank-constrained SCM estimation as

$$\mathcal{L}(r_{ij}^{(h)}, r_{ij}^{(u)}, \lambda_i) = \sum_{i,j} \Bigg[\boldsymbol{x}_{ij}^{\mathrm{H}} (\mathbf{R}_{ij}^{(x)})^{-1} \boldsymbol{x}_{ij} + \log\det \mathbf{R}_{ij}^{(x)} + (\alpha+1)\log r_{ij}^{(h)} + \frac{\beta}{r_{ij}^{(h)}} \Bigg] + \text{const.}, \tag{14}$$

where const. does not depend on the objective variables. The parameters of this negative log likelihood function $\mathcal{L}$ are optimized by a maximum a posteriori estimation based on the expectation–maximization (EM) algorithm [11].

The Q function is defined by the expected complete-data log-likelihood values regarding the a posteriori probability $p(s_{ij}^{(h)}, \boldsymbol{u}_{ij} \mid \boldsymbol{x}_{ij}; \tilde{\Theta})$ as

$$Q(\Theta; \tilde{\Theta}) = \sum_{i,j} \Bigg[-(\alpha+2)\log r_{ij}^{(h)} - M \log r_{ij}^{(u)} - \log\det \mathbf{R}_i^{(u)} - \frac{\hat{r}_{ij}^{(h)} + \beta}{r_{ij}^{(h)}} - \frac{\mathrm{tr}\left(\left(\mathbf{R}_i^{(u)} \right)^{-1} \hat{\mathbf{R}}_{ij}^{(u)} \right)}{r_{ij}^{(u)}} \Bigg] + \text{const.}, \tag{15}$$

where $\Theta = \{r_{ij}^{(h)}, r_{ij}^{(u)}, \lambda_i\}$ is set of the set of parameters to be updated, $\tilde{\Theta} = \{\tilde{r}_{ij}^{(h)}, \tilde{r}_{ij}^{(u)}, \tilde{\lambda}_i\}$ is the up-to-date parameters, and $\hat{r}_{ij}^{(h)}$ and $\hat{\mathbf{R}}_i^{(u)}$ are the sufficient statistics obtained by the E-step. The update rules in the E-step are expressed as

$$\tilde{\mathbf{R}}_i^{(u)} = \mathbf{R}'^{(u)}_i + \tilde{\lambda}_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{H}}, \tag{16}$$

$$\mathbf{R}_{ij}^{(x)} = \tilde{r}_{ij}^{(h)} \boldsymbol{a}_i^{(h)} (\boldsymbol{a}_i^{(h)})^{\mathrm{H}} + \tilde{r}_{ij}^{(u)} \tilde{\mathbf{R}}_i^{(u)}, \tag{17}$$

$$\hat{r}_{ij}^{(h)} = \tilde{r}_{ij}^{(h)} - \left(\tilde{r}_{ij}^{(h)} \right)^2 \left(\boldsymbol{a}_i^{(h)} \right)^{\mathrm{H}} \left(\mathbf{R}_{ij}^{(x)} \right)^{-1} \boldsymbol{a}_i^{(h)} + \left| \tilde{r}_{ij}^{(h)} \boldsymbol{x}_{ij}^{\mathrm{H}} \left(\mathbf{R}_{ij}^{(x)} \right)^{-1} \boldsymbol{a}_i^{(h)} \right|^2, \tag{18}$$

$$\hat{\mathbf{R}}_{ij}^{(u)} = \tilde{r}_{ij}^{(u)} \tilde{\mathbf{R}}_i^{(u)} - \left(\tilde{r}_{ij}^{(u)} \right)^2 \tilde{\mathbf{R}}_i^{(u)} \left(\mathbf{R}_{ij}^{(x)} \right)^{-1} \tilde{\mathbf{R}}_i^{(u)} + \left(\tilde{r}_{ij}^{(u)} \right)^2 \tilde{\mathbf{R}}_i^{(u)} \left(\mathbf{R}_{ij}^{(x)} \right)^{-1} \boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathrm{H}} \left(\mathbf{R}_{ij}^{(x)} \right)^{-1} \tilde{\mathbf{R}}_i^{(u)}. \tag{19}$$

In the M-step, a coordinate ascent algorithm is applied to the Q function, i.e.,

$$r_{ij}^{(h)} \leftarrow \frac{\hat{r}_{ij}^{(h)} + \beta}{\alpha + 2}, \tag{20}$$

$$\mathbf{K}_i = \frac{1}{J} \sum_j \frac{1}{\tilde{r}_{ij}^{(u)}} \hat{\mathbf{R}}_{ij}^{(u)}, \tag{21}$$

$$\lambda_i = \boldsymbol{b}_i^{\mathrm{H}} \mathbf{K}_i \boldsymbol{b}_i, \tag{22}$$

$$\mathbf{R}_i^{(u)} \leftarrow \mathbf{R}'^{(u)}_i + \lambda_i \boldsymbol{b}_i \boldsymbol{b}_i^{\mathrm{H}}, \tag{23}$$

$$r_{ij}^{(u)} \leftarrow \frac{1}{M} \mathrm{tr}\left(\left(\mathbf{R}_i^{(u)} \right)^{-1} \hat{\mathbf{R}}_{ij}^{(u)} \right). \tag{24}$$

TABLE I
RECORDING CONDITIONS

Recording location	Studio
Reverberation time (T_{60})	300 ms
Microphone	C417 PP (AKG)
Loudspeaker	ADIVA11 (Anthony Gallo)
Microphone preamplifier	Octamic II (RME)
Audio interface	828x (MOTU)
TSP length	65536 samples
Recording sampling freq.	48 kHz

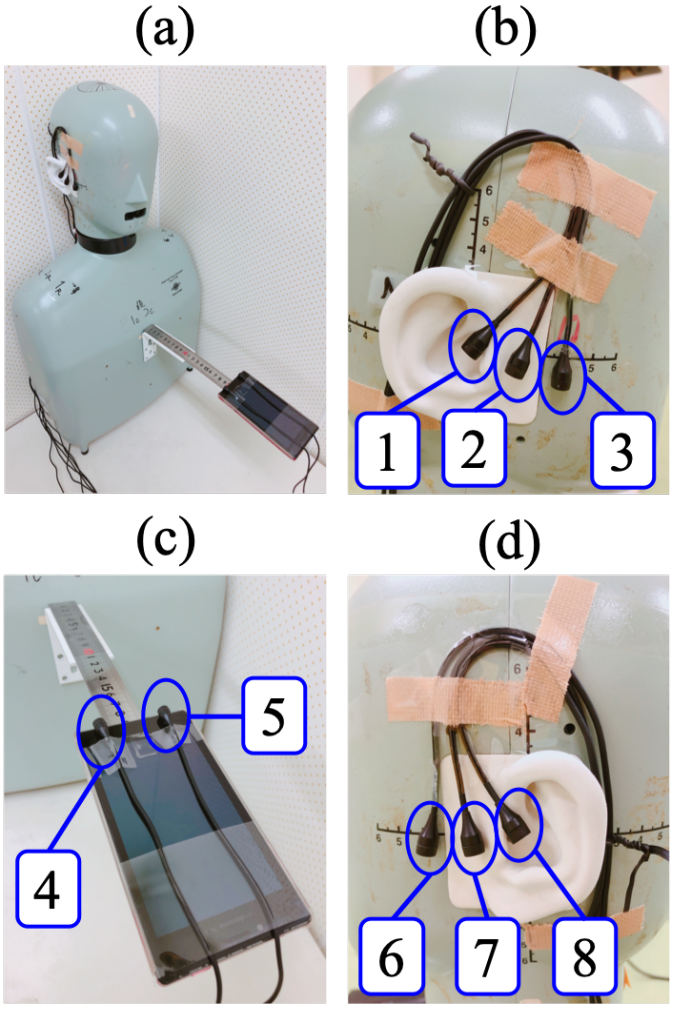

Fig. 1. (a) Overall view of head-and-torso dummy, (b) right-ear microphone array, (c) smartphone microphones, and (d) left-ear microphone array.

After the convergence of the EM algorithm, we can construct the multichannel Wiener filter [11] using the estimated $r_{ij}^{(h)}$, $r_{ij}^{(u)}$, and $\mathbf{R}_i^{(u)}$ with λ_i, and output the extracted target speech component.

III. Multichannel Hearing Aid System

A. Specification of Apparatus

In this experimental scheme we assume the situation where a person wearing a multichannel hearing aid talks with someone facing him/her. We constructed the recording system and recorded impulse responses and diffuse noise using eight microphones that are synchronized with the same sampling rate. Figure 1(a) shows the head-and-torso dummy, which imitates a person wearing a binaural hearing aid and holding a smartphone. Three microphones are attached to each ear [see Figs. 1(b) and (d)]. The smartphone is attached 20 cm apart from the chest and two microphones with the interelement spacing of 4 cm are set on the smartphone [see Fig. 1(c)]. For convenience, we number each microphone as shown in Figs. 1(b)–(d). The height of the dummy head is set to 170 cm and the loudspeaker is set in front of the dummy head to mimic the situation of conversation. Accordingly, the height of the loudspeaker is set at 152 cm, corresponding to the mouth position of the conversation partner whose head is of the same height as the dummy head.

B. Recording of Impulse Response and Diffuse Noise

The time stretched pulse (TSP) signal is adopted in the measurement. The conditions are shown in Table I. Eight microphones are synchronized by the audio interface in this research. When the proposed hearing aid system is used in an actual situation, it is necessary to apply the following methods to synchronize these microphones [16], [17]. The distance from the dummy head to the loudspeaker is varied by 75, 100, and 150 cm, and the angle is varied by -20, 0, and $20°$, where $0°$ means the normal to the dummy head. The overview of the positions in nine recordings is shown in Fig. 2.

The diffuse noise is supposed to be a sound in a crowded place, where many people talk and walk freely. We instructed the volunteer subjects to walk around the dummy head and read the designated sentences aloud. Assuming that the noise sources surround the target source, we instructed them to walk outside the 150 cm radius of the front half circle of the dummy head.

IV. Experimental Evaluation

A. Experimental Conditions

The purpose of this experiment is to evaluate the applicability of ILRMA and the rank-constrained SCM estimation for a multichannel hearing aid system in a real environment. Female utterance was convolved with the impulse response to produce the target speech. We used the utterances from the JNAS database [18] and the recorded impulse response described in Sec. III-B. Since the sampling rate of the corpus is 16 kHz, the impulse response and diffuse noise were down-sampled from 48 to 16 kHz. The observed signal was generated by mixing the recorded diffuse noise and the target signal at the input SNRs of -10, -5, and 0 dB. In ILRMA, the observed signal was preprocessed by sphering transformation by principal component analysis. In the rank-constrained SCM estimation, the shape parameter α was experimentally chosen and set to $0.5, 1.1, 10$, and 20, and the scale parameter β was set to 10^{-16}. Initialization trials were conducted ten times using different random values. The other conditions are shown in Table II. We used source-to-distortion ratio (SDR) improvement [19] as the objective measurement citation.

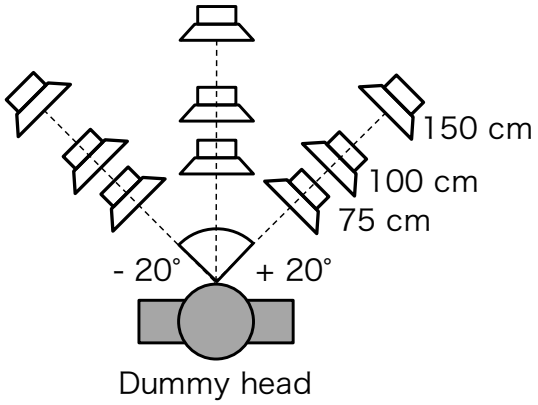

Fig. 2. Position of loudspeaker (mouth of conversation partner) for nine recordings.

TABLE II
Experimental conditions for BSS

Sampling freq.	16 kHz
FFT length	1024 sample (50% overlap)
Window	Hamming window
Number of bases in low-rank model	10
Number of iterations in ILRMA	50
Initialization of $\boldsymbol{W}_i$ in ILRMA	Identity matrix
Number of iterations in rank-constrained SCM estimation	10

B. SDR Improvement Behavior

Figure 3 shows that the average SDR tended to improve for every iteration under the -10 dB input SNR condition at microphone 1 (nearest right external auditory canal). From the figure, we can confirm that the rank-constrained SCM estimation outperforms ILRMA in all recordings. The value of the hyperparameter α greatly affects the SDR improvement and it is better to set α large. Moreover, the rank-constrained SCM estimation can achieve the highest SDR improvement by the second or third iteration, showing the advantage of the fast convergence.

C. SDR Improvement at Various Input SNRs

We concentrate our attention on the target speech angle of $0°$ and investigate SDR improvements by ILRMA and rank-constrained SCM estimation in each input SNR. Note that the second iteration score of the the rank-constrained SCM estimation was adopted on the basis of the results in Sec. IV-B.

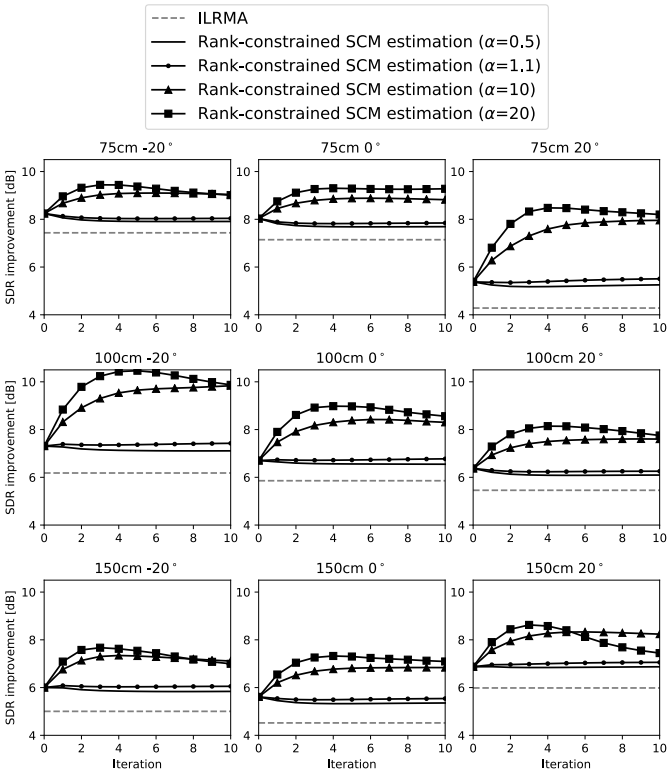

Fig. 3. Average SDR improvements for each iteration at microphone 1 under -10 dB input SNR condition.

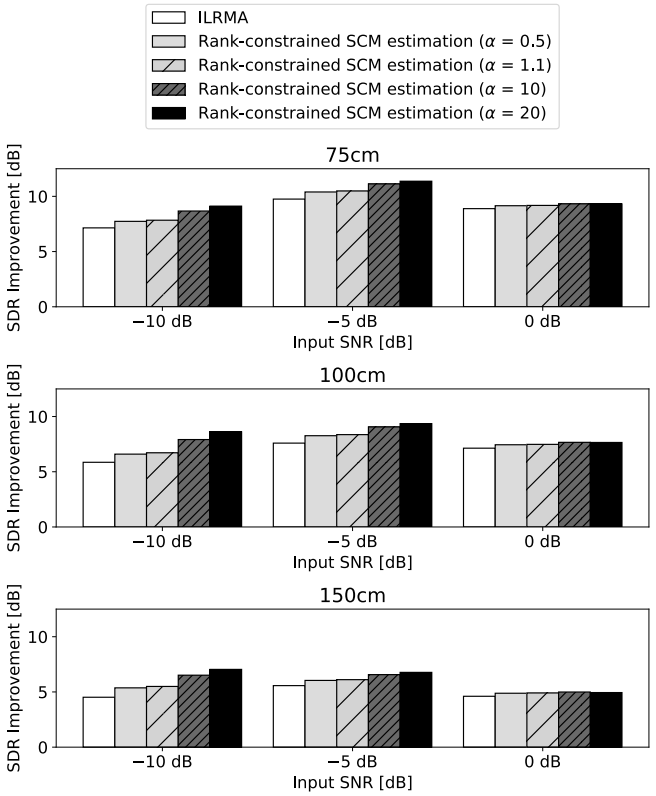

Fig. 4. Average SDR improvements of ILRMA and rank-constrained SCM estimation after two iterations at microphone 1 when target source is located at 0°.

Figure 4 shows the average SDR improvement at microphone 1 in each input SNR. From Fig. 4, ILRMA achieves acceptable separation performance compared with the observed signal. The SDR improvements by the rank-constrained SCM estimation are higher than those by ILRMA especially at lower input SNRs (i.e., -10 and -5 dB). Furthermore, we can confirm that the sparse prior corresponding to a larger α improves the separation performance in all the recordings.

V. Conclusions

In this study, we investigate the applicability of ILRMA and the rank-constrained SCM estimation with a binaural hearing aid system including microphones on a smartphone in a real environment. We construct the experimental system to record the impulse response and diffuse noise. Using the recorded data, we evaluate the separation performance of ILRMA and the rank-constrained SCM estimation . The experimental results show that ILRMA and the rank-constrained SCM estimation work well for a the binaural hearing aid scheme. Furthermore, the separation performance of the rank-constrained SCM estimation is higher than that of ILRMA especially under low-SNR conditions.

Acknowledgment

This work was partly supported by SECOM Science and Technology Foundation and JSPS KAKENHI Grant Numbers JP19H01116 and 19K20306.

References

[1] 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 Trans. Signal and Information Processing*, vol. 8, no. e12, pp. 1–14, 2019.

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

[3] 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.

[4] A. Hiroe, "Solution of permutation problem in frequency domain ICA using multivariate probability density functions," in *Proc. ICA*, 2006, pp. 601–608.

[5] 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.

[6] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 189–192, 2011.

[7] 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 Trans. ASLP*, vol. 24, no. 9, pp. 1626–1641, 2016.

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

[9] D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, and K. Kondo, "Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source

separation," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 28, pp. 1–25, 2018.

[10] 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.

[11] Y. Kubo, N. Takamune, D. Kitamura, and H. Saruwatari, "Efficient full-rank spatial covariance estimation using independent low-rank matrix analysis for blind source separation," in *Proc. EUSIPCO*, 2019.

[12] N. Q. K. Duong, E. Vincent, and R. Gribonval, "Underdetermined reverberant audio source separation using a full-rank spatial covariance model," *IEEE Trans. ASLP*, vol. 18, no. 7, pp. 1830–1840, 2010.

[13] A. Ozerov and C. Févotte, "Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation," *IEEE Trans. ASLP*, vol. 18, no. 3, pp. 550–563, 2010.

[14] 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.

[15] Y. Takahashi, T. Takatani, K. Osako, H. Saruwatari, and K. Shikano, "Blind spatial subtraction array for speech enhancement in noisy environment," *IEEE Trans. ASLP*, vol. 17, no. 4, pp. 650–664, 2009.

[16] S. Miyabe, N. Ono, and S. Makino, "Blind compensation of interchannel sampling frequency mismatch for ad hoc microphone array based on maximum likelihood estimation," *Elsevier Signal Processing*, vol. 107, no. 2015, pp. 185–196, 2015.

[17] R. Sakanashi, N. Ono, S. Miyabe, T. Yamada, and S Makino, "Speech enhancement with ad-hoc microphone array using single source activity," in *Proc. APSIPA2013*, 2013, pp. 1–6.

[18] K. Itou, M. Yamamoto, K. Takeda, T. Takezawa, T. Matsuoka, T. Kobayashi, K. Shikano, and S. Itahashi, "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *The Journal of Acoustical Society of Japan (E)*, vol. 20, no. 3, pp. 199–206, 1999.

[19] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.