Ego Noise Reduction for Hose-Shaped Rescue Robot Combining Independent Low-Rank Matrix Analysis and Multichannel Noise Cancellation

Narumi Mae[1(✉)], Masaru Ishimura[1], Shoji Makino[1], Daichi Kitamura[2], Nobutaka Ono[2,3], Takeshi Yamada[1], and Hiroshi Saruwatari[4]

[1] University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki 305-8573, Japan
{mae,ishimura}@mmlab.cs.tsukuba.ac.jp, maki@tara.tsukuba.ac.jp, takeshi@cs.tsukuba.ac.jp

[2] SOKENDAI (The Graduate University for Advanced Studies), Shonan Village, Hayama, Kanagawa 240-0193, Japan
d-kitamura@nii.ac.jp

[3] National Institute of Informatics (NII), 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan
onono@nii.ac.jp

[4] The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan
hiroshi_saruwatari@ipc.i.u-tokyo.ac.jp

Abstract. In this paper, we present an ego noise reduction method for a hose-shaped rescue robot, developed for search and rescue operations in large-scale disasters. It is used to search for victims in disaster sites by capturing their voices with its microphone array. However, ego noises are mixed with voices, and it is difficult to differentiate them from a call for help from a disaster victim. To solve this problem, we here propose a two-step noise reduction method involving the following: (1) the estimation of both speech and ego noise signals from observed multichannel signals by multichannel nonnegative matrix factorization (NMF) with the rank-1 spatial constraint, and (2) the application of multichannel noise cancellation to the estimated speech signal using reference signals. Our evaluations show that this approach is effective for suppressing ego noise.

Keywords: Rescue robot · Tough environment · Noise reduction · Nonnegative matrix factorization · Independent vector analysis · Multichannel noise cancellation

1 Introduction

It is important to develop robots for search and rescue operations during large-scale disasters such as earthquakes. Robots are required for emergency responses and for the restoration of disaster sites, which are difficult and dangerous tasks

P. Tichavský et al. (Eds.): LVA/ICA 2017, LNCS 10169, pp. 141–151, 2017.
DOI: 10.1007/978-3-319-53547-0_14

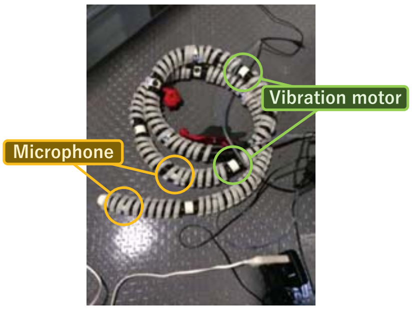

Fig. 1. Hose-shaped rescue robot.

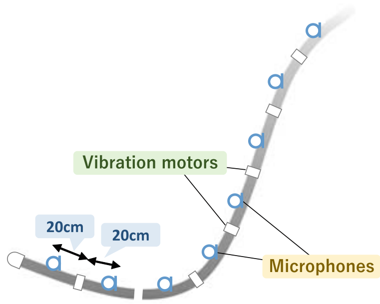

Fig. 2. Structure of hose-shaped rescue robot.

for humans. The "Tough Robotics Challenge" is one of the research and development programs in the Impulsing Paradigm Change through Disruptive Technologies Program (ImPACT) [1]. One of the robots developed in this program is a hose-shaped rescue robot [2]. This robot is long and slim like a snake, allowing it to investigate narrow spaces into which conventional remotely operable robots cannot enter. This robot searches for a disaster victim by capturing his/her voice using a microphone array attached around itself at regular intervals. However, there is a serious problem of "ego noise". This noise is generated by the vibration motors used to move the robot via the vibrating cilia tape wrapped around the robot. In this study, we focus on reducing the ego noise from the sound recorded by the microphone array of the robot.

Recently, many ego noise reduction methods have been proposed [3–6]. In addition, the many microphones on the hose-shaped rescue robot enable the application of the overdetermined source separation method. However, the microphone arrangement changes as the robot moves, making it difficult to control the microphone array geometry. Hence, in [7], we proposed a noise reduction method for the hose-shaped rescue robot combining determined rank-1 multichannel nonnegative matrix factorization [8,9] proposed by Kitamura *et al.* which can be interpreted as an independent low-rank matrix analysis (hereafter referred to as ILRMA), and a noise canceller (NC). As a reference input of the NC, we used the sum of all the noise components of the ILRMA outputs. On the other hand, in this study, we use a multichannel NC and confirm the applicability of the proposed method for reducing ego noise.

2 Hose-Shaped Rescue Robot and Ego Noise

2.1 Hose-Shaped Rescue Robot

Figure 1 shows an image of the hose-shaped rescue robot and Fig. 2 shows its structure. The hose-shaped rescue robot basically consists of a hose as its axis

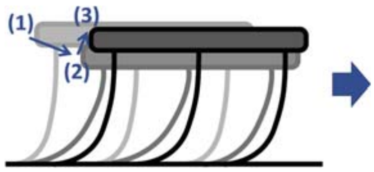

Fig. 3. Principle of movement of hose-shaped rescue robot [2].

with cilia tape wrapped around it; it moves forward slowly as a result of the reaction between the cilia and floor through the vibration of the cilia tape with the vibration motors. Figure 3 schematically shows the principle of movement of the hose-shaped rescue robot. When the motors vibrate, state (1) changes to state (2) through the friction between the cilia and the floor, then state (2) changes to state (3) as a result of the cilia slipping. The hose-shaped rescue robot moves by repeating such changes in its state. It performs various sensing functions using sensors such as microphones, cameras, an inertial measurement unit and light sensors.

2.2 Problem in Recording Speech

Recording speech using the hose-shaped rescue robot has a serious problem. During the operation of the robot, very loud ego noise is mixed in the input to the microphones. The main sources of the ego noise are the driving sound of the vibration motors, the fricative sound generated between the cilia and floor, and the noise generated by microphone vibration. In an actual disaster site, the voice of a person seeking help is not sufficiently loud to capture and it is smaller than the ego noise.

3 Overview of Independent Low-Rank Matrix Analysis

Recently, many ego noise reduction methods have been proposed [3–6]. In [3], noise reduction based on generalizations of K-singular value decomposition (K-SVD) was proposed, which can be used for an underdetermined multichannel situation. Also, the authors of [4,5] proposed a method of improving the performance of ego noise reduction using an adaptive microphone array geometry. On the other hand, the many microphones on the rescue robot enable the application of an overdetermined source separation method. In a determined situation, independent vector analysis (IVA) [10–12] is a commonly used method. IVA requires independence between the sources to estimate a demixing matrix. In general, in IVA, a spherical multivariate distribution is assumed as the source model to ensure a higher-order correlation between the frequency bins in all sources. However, this model does not include any particular information on the sources, that is, IVA cannot capture specific spectral structures of the sources. Thus, the utilization of nonnegative matrix factorization (NMF) [13–15] as the

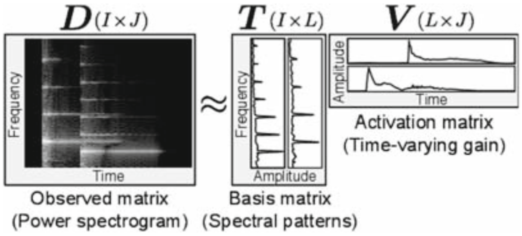

Fig. 4. Decomposition model of NMF.

source model has been proposed [8,9], which enables us to capture the spectral structures.

NMF decomposes a given spectrogram into several spectral bases $\boldsymbol{T}$ and temporal activations $\boldsymbol{V}$, as shown in Fig. 4, then the decomposed components are clustered into each source. Multichannel NMF (MNMF) [16–18] is one of the techniques for clustering the NMF bases and activations using a sourcewise spatial model. MNMF separately models the mixing system and the nonnegative power spectra of sources. However, this method is strongly dependent on its initial values because there are no constraints in the spatial models.

To solve the problem of MNMF, ILRMA [8,9] was proposed, in which a rank-1 spatial model is introduced into MNMF [18]. This method estimates a demixing matrix while representing a source using NMF bases, and can be optimized by the update rules of IVA and conventional single-channel NMF. Therefore, ILRMA is a method that unifies IVA and NMF.

Since the hose-shaped rescue robot moves very slowly and the spatial locations of the sources and microphones barely change, we can assume a linear time-varying mixing system. In this case, ILRMA is effective for the separation because it does not require the locations of the sources and microphones. In particular, ILRMA can efficiently capture the time-frequency structure of the ego noise because it is the repetition of several types of similar spectra.

3.1 Formulation

We assume that M sources are observed using M microphones (determined case). The sources, the observed and separated signals in each time-frequency slot are as follows:

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

where $1 \leq i \leq I$ and $1 \leq j \leq J$ are indexes of frequency and time, respectively, and t denotes the vector transpose. All the entries of these vectors are complex values. When the window size in an STFT is sufficiently longer than the impulse response between a source and microphone, we can approximately represent the observed signal as

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

Here, $\boldsymbol{A}_i = (\boldsymbol{a}_{i,1} \ \cdots \ \boldsymbol{a}_{i,M})$ is an $M \times M$ mixing matrix of the observed signals. When $\boldsymbol{W}_i = (\boldsymbol{w}_{i,1} \ \cdots \ \boldsymbol{w}_{i,M})^h$ denotes the demixing matrix, the separated signal $\boldsymbol{y}_{ij}$ is represented as

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

where ^{h} is the Hermitian transpose.

3.2 Independent Low-Rank Matrix Analysis

We use ILRMA [8,9] to impose a rank-1 spatial model on MNMF [18]. We explain the formulation and algorithm derived by Kitamura *et al.* [8,9] MNMF is an extension of simple NMF for multichannel signals. The observed signals are represented as

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

where X_{ij} of size $M \times M$ is the correlation matrix between channels. The diagonal elements of X_{ij} represent real-valued powers detected by the microphones, and the nondiagonal elements represent the complex-valued correlations between the microphones. The separation model of MNMF $\hat{\mathbf{X}}_{ij}$ used to approximate $\mathbf{X}_{ij}$ is represented as

$$\mathbf{X}_{ij} \approx \hat{\mathbf{X}}_{ij} = \sum_m \mathbf{H}_{i,m} \sum_l t_{il,m} v_{lj,m}, \tag{7}$$

where $m = 1 \cdots M$ is the index of the sound sources. $\mathsf{H}_{i,m}$ is an $M \times M$ spatial covariance matrix for each frequency i and source m, and $\mathsf{H}_{i,m} = \boldsymbol{a}_{i,m} \boldsymbol{a}_{i,m}^h$ is limited to a rank-1 matrix. This assumption corresponds to $t_{il,m} \in \mathbb{R}_+$ and $v_{lj,m} \in \mathbb{R}_+$ being the elements of the basis matrix $\boldsymbol{T}_m$ and activation matrix $\boldsymbol{V}_m$, respectively. This rank-1 spatial constraint leads to the following cost function:

$$\mathcal{Q} = \sum_{i,j} \left[\sum_m \frac{|y_{ij,m}|^2}{\sum_l t_{il,m} v_{lj,m}} - 2 \log |\det \boldsymbol{W}_i| + \sum_m \log \sum_l t_{il,m} v_{lj,m} \right], \tag{8}$$

namely, the estimation of $\mathsf{H}_{i,m}$ can be transformed to the estimation of the demixing matrix $\boldsymbol{W}_i$. This cost function is equivalent to the Itakura-Saito divergence between X_{ij} and $\hat{\mathsf{X}}_{ij}$, and we can derive

$$t_{il,m} \leftarrow \quad t_{il,m}\sqrt{\frac{\Sigma_j |y_{ij,m}|^2 v_{lj,m} \left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-2}}{\Sigma_j v_{lj,m} \left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-1}}}, \tag{9}$$

$$v_{lj,m} \leftarrow \quad v_{il,m}\sqrt{\frac{\Sigma_i |y_{ij,m}|^2 t_{il,m} \left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-2}}{\Sigma_i t_{il,m} \left(\Sigma_{l'} t_{il',m} v_{l'j,m}\right)^{-1}}}, \tag{10}$$

$$r_{ij,m} = \quad \sum_l t_{il,m} v_{lj,m}, \tag{11}$$

$$V_{i,m} = \quad \frac{1}{J}\sum_j \frac{1}{r_{ij,m}} \boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^h, \tag{12}$$

$$\boldsymbol{w}_{i,m} \leftarrow \quad \left(\boldsymbol{W}_i V_{i,m}\right)^{-1} \boldsymbol{e}_m, \tag{13}$$

where $\boldsymbol{e}_m$ is a unit vector whose mth element is one. We can simultaneously estimate both the sourcewise time-frequency model $r_{ij,m}$ and the demixing matrix $\boldsymbol{W}_i$ by iterating (9)–(13) alternately. After the cost function converges, the separated signal $\boldsymbol{y}_{ij}$ can be obtained as (5). Note that since the signal scale of $\boldsymbol{y}_{ij}$ cannot be determined, we apply a projection-back method [19] to $\boldsymbol{y}_{ij}$ to determine the scale.

The demixing filter in ILRMA is time-invariant over several seconds. To achieve time-variant noise reduction, in a previous study [7], we applied a single-channel NC for the postprocessing of ILRMA to reduce the remaining time-variant ego noise components. An NC usually requires a reference microphone to observe only the noise signal. Thus, we utilized the noise estimates obtained by ILRMA as the noise reference signals.

4 Multichannel Noise Canceller

4.1 Conventional Method

The NC proposed in [7,20] requires a reference microphone located near a noise source. The recorded noise reference signals $n_1(t), \ldots, n_k(t)$ are utilized to reduce

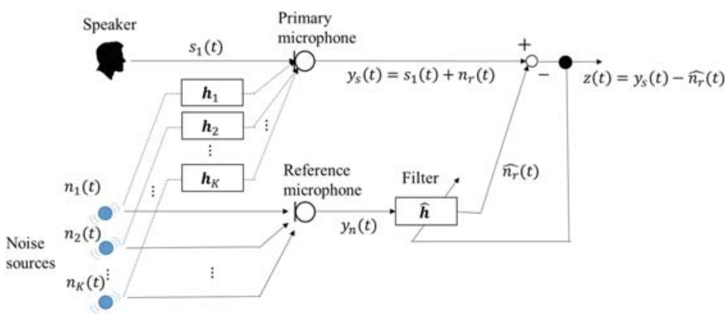

Fig. 5. Noise canceller.

the noise in the observed speech signal $s_1(t)$ as shown in Fig. 5. We here assume that both $s_1(t)$ and $n_1(t), \ldots, n_k(t)$ are simultaneously recorded. The observed signal contaminated with the noise source can be represented as

$$y_s(t) = s_1(t) + n_1(t) + \cdots + n_k(t). \tag{14}$$

We can consider that the noise signal $n_r(t)$ is strongly correlated with the reference noise signal $y_n(t)$ and that $n_r(t) = n_1(t) + \cdots + n_k(t)$ can be represented by a linear convolution model as

$$n_r(t) \simeq \hat{n}_r(t) = \hat{\boldsymbol{h}}(t)^t \boldsymbol{y}_n(t), \tag{15}$$

where $\boldsymbol{y}_n(t) = [y_n(t)\; y_n(t-1)\; \cdots\; y_n(t-N+1)]^t$ is the reference microphone input from the current time t to the past N samples and $\hat{\boldsymbol{h}}(t) = [\hat{\boldsymbol{h}}_1(t)\; \hat{\boldsymbol{h}}_2(t) \cdots \hat{\boldsymbol{h}}_N(t)]^t$ is the estimated impulse response. From (15), the speech signal $s_1(t)$ is extracted by subtracting the estimated noise $\hat{\boldsymbol{h}}(t)^t \boldsymbol{y}_n(t)$ from the observation as

$$z(t) = x(t) - \hat{\boldsymbol{h}}(t)^t \boldsymbol{y}_n(t), \tag{16}$$

where $z(t)$ is the estimated speech signal.

4.2 Proposed Method

In the conventional NC proposed in [7,20], we used the sum of all the noise components of the ILRMA outputs applied projection-back method to the same microphone. This means that the change in each mixing system $\boldsymbol{h_1}, \boldsymbol{h_2}, \ldots, \boldsymbol{h_k}$ in Fig. 5 is the same. However, the mixture systems may change differently according to the noise sources. Thus, in this study, we use a multichannel NC. Figure 6 shows the multichannel NC model. In this model, the filter of the multichannel NC is estimated for each noise source. Thereby, the filter and noise can be more precisely estimated. The filter $\hat{\boldsymbol{h}}(t)$ can be obtained by minimization of the mean square error. In this paper, we use the normalized least mean square (NLMS) algorithm [21] to estimate $\hat{\boldsymbol{h}}(t)$. From the NLMS algorithm, the update rule of the filter $\hat{\boldsymbol{h}}(t)$ is given as

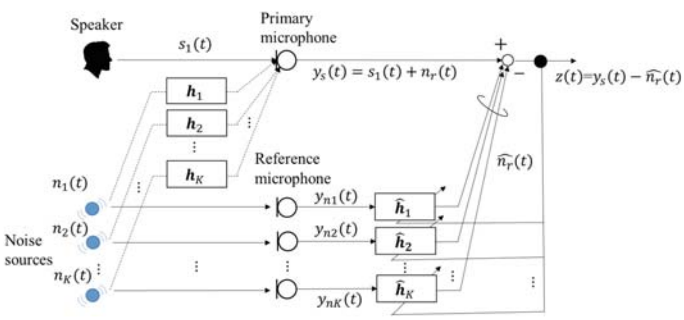

Fig. 6. Multichannel noise canceller.

$$\hat{\boldsymbol{h}}(t+1) = \hat{\boldsymbol{h}}(t) + \mu \frac{z(t)}{||\boldsymbol{y}_n(t)||^2} \boldsymbol{y}_n(t), \tag{17}$$

where

$$\hat{\boldsymbol{h}}(t) = [\hat{\boldsymbol{h}}_1(t)^t \ \hat{\boldsymbol{h}}_2(t)^t \cdots \ \hat{\boldsymbol{h}}_K(t)^t]^t, \tag{18}$$

$$\hat{\boldsymbol{h}}_k(t) = [\hat{h}_{k,0}(t) \ \hat{h}_{k,1}(t) \cdots \ \hat{h}_{k,N-1}(t)]^t, \tag{19}$$

$$\boldsymbol{y}_n(t) = [\boldsymbol{y}_{n1}(t)^t \ \boldsymbol{y}_{n2}(t)^t \cdots \ \boldsymbol{y}_{nK}(t)^t]^t, \tag{20}$$

$$\boldsymbol{y}_{nk}(t) = [y_{nk}(t), y_{nk}(t-1), \cdots, y_{nk}(t-N+1)]^t. \tag{21}$$

4.3 Flow of the Proposed Method

Figure 7 shows the flow of the proposed method. In Fig. 7, $y_{1(t)}$, ..., $y_{8(t)}$ are the ILRMA outputs, $y_s(t)$ is the speech signal estimated by ILRMA, and $y_{n1(t)}$, ..., $y_{n7(t)}$ are the residual outputs corresponding to the various components of ego noise. In the first step, the observed signals are separated into independent signals via ILRMA, where the number of separated signals is the same as the number of microphones ($M = 8$). Note that ILRMA cannot determine the order of the output signals. Therefore, we must find an estimated signal that includes most of the speech components to be used as $y_s(t)$. In this paper, we manually choose the speech estimate from the output signals, while such speech estimate detection may be possible by employing statistics or spectrograms of the output signals. Since a time-invariant spatial demixing matrix (demixing filter) is applied for the separation in the first step, the ego noise, which does not follow the time-invariant assumption, remains in the separated speech signal $y_s(t)$. In the second step, we apply the multichannel NC with the ego noise reference signals $y_{n1(t)}$, ..., $y_{n7(t)}$. In this step, we expect that the multichannel NC will reduce the residual noise component in $y_s(t)$ because the multichannel NC models the time-variant noise as $\hat{\boldsymbol{h_1}}(t)^t \boldsymbol{y}_{n1}(t)$, ..., $\hat{\boldsymbol{h_7}}(t)^t \boldsymbol{y}_{n7}(t)$, which can update the filter $\hat{\boldsymbol{h_1}}(t), \ldots, \hat{\boldsymbol{h_7}}(t)$ for each time sample.

5 Experiment

5.1 Conditions

In this experiment, we measured an observed signal using the hose-shaped rescue robot. This robot consists of eight microphones and seven vibration motors, and its total length is approximately 3 m. The recorded speech signal was produced by convolving a dry speech signal and the measured impulse responses between a disaster victim and the microphones on the robot. For the noise signal, we recorded actual ego noise by moving the robot in an area that simulated a disaster site. The observed multichannel signals were obtained as the sum of these speech and ego noise signals in each microphone, namely, they were a mixture of time-invariant speech and time-variant actual ego noises. In addition, we compared three methods: simple ILRMA, ILRMA with a single-channel NC,

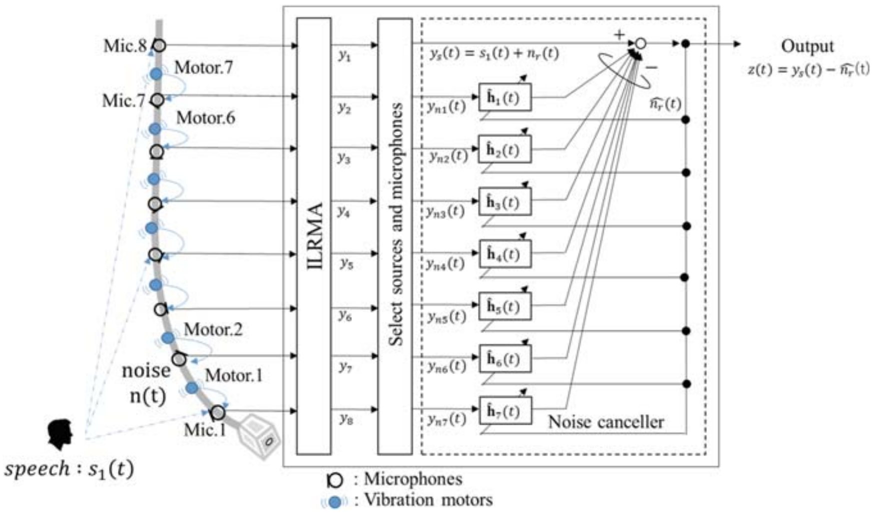

Fig. 7. Flow of proposed method.

Table 1. Experimental conditions

Sampling frequency	16 kHz
Window length	1024 samples
Window shift	STFT length/4
Number of bases	15
Number of iterations	100
Filter length of noise canceller	1600 taps
Step size of NLMS	0.1
Input SNR	0, −5, −10 dB

and the proposed method (ILRMA + multichannel NC). The signal-to-distortion ratio (SDR) and the signal-to-interference ratio (SIR) [22] were used to evaluate the separation performance. The other experimental conditions are shown in Table 1. The estimated signal that includes most of the speech components was projected back to microphone 1. Also, each estimated noise signal was projected back to microphone 2.

5.2 Results

Figure 8 shows the improvements in the SDR and SIR for each method. The results show that the multichannel NC improves the separation performance. This is because the multichannel NC efficiently estimates the changes in each filter from the estimation result of ILRMA.

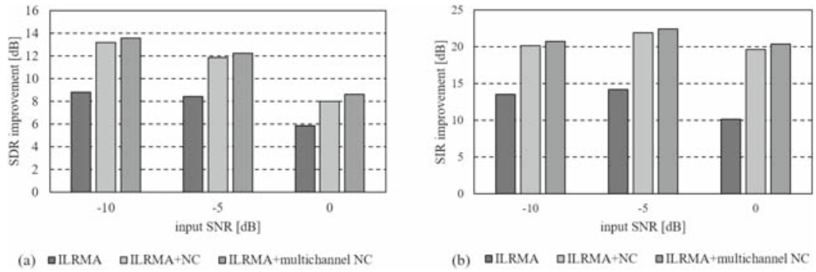

Fig. 8. (a) SDR and (b) SIR improvements for recording at SNRs = −10, −5 and 0 dB.

6 Conclusion

To enhance speech signals recorded by a hose-shaped rescue robot, we have proposed an ego noise reduction method using ILRMA and multichannel NC. We evaluated the proposed method by an experiment and compared ILRMA, ILRMA with a single-channel NC, and ILRMA with a multichannel NC in terms of the SDR and SIR. It was found that the proposed method exhibited the best performance under all conditions, thus confirming the effectiveness of combining ILRMA and the multichannel NC.

Acknowledgments. This work was supported by the Japan Science and Technology Agency and the Impulsing Paradigm Change through Disruptive Technologies Program (ImPACT) designed by the Council for Science, Technology and Innovation, and partly supported by SECOM Science and Technology Foundation. We would also like to express our gratitude to Prof. Hiroshi Okuno and Mr. Yoshiaki Bando for providing experimental data.

References

1. Impulsive Paradigm Change through Disruptive Technologies Program (ImPACT). http://www.jst.go.jp/impact/program07.html
2. Namari, H., Wakana, K., Ishikura, M., Konyo, M., Tadokoro, S.: Tube-type active scope camera with high mobility and practical functionality. In: Proceedings of IEEE/RSJ IROS, pp. 3679–3686 (2012)
3. Deleforge, A., Kellerman, W.: Phase-optimized K-SVD for signal extraction from underdetermined multichannel sparse mixtures. In: Proceedings of IEEE ICASSP, pp. 355–359 (2015)
4. Barfuss, H., Kellerman, W.: Improving blind source separation performance by adaptive array geometries for humanoid robots. In: Proceedings of HSCMA (2014)
5. Barfuss, H., Kellerman, W.: An adaptive microphone array topology for target signal extraction with humanoid robots. In: Proceedings of IWAENC, pp. 16–20 (2014)
6. Aichner, R., Zourub, M., Buchner, H., Kellerman, W.: Post-processing for convolutive blind source separation. In: Proceedings of ICASSP (2006)

7. Mae, N., Kitamura, D., Ishimura, M., Yamada, T., Makino, S.: Ego noise reduction for hose-shaped rescue robot combining independent low-rank matrix analysis and noise cancellation. In: Proceedings of APSIPA (2016, to be published)
8. Kitamura, D., Ono, N., Sawada, H., Kameoka, H., Saruwatari, H.: Efficient multichannel nonnegative matrix factorization exploiting rank-1 spatial model. In: Proceedings of ICASSP, pp. 276–280 (2015)
9. Kitamura, D., Ono, N., Sawada, H., Kameoka, H., Saruwatari, H.: Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. IEEE/ACM Trans. Audio Speech Lang. Process. **24**(9), 1626–1641 (2016)
10. Kim, T., Eltoft, T., Lee, T.-W.: Independent vector analysis: an extension of ICA to multivariate components. In: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 165–172. Springer, Heidelberg (2006). doi:10.1007/11679363_21
11. Hiroe, A.: Solution of permutation problem in frequency domain ICA, using multivariate probability density functions. In: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) ICA 2006. LNCS, vol. 3889, pp. 601–608. Springer, Heidelberg (2006). doi:10.1007/11679363_75
12. Kim, T., Attias, H.T., Lee, S.-Y., Lee, T.-W.: Blind source separation exploiting higher-order frequency dependencies. IEEE Trans. Speech Audio Process. **15**(1), 70–79 (2007)
13. Lee, D.D., Seung, H.S.: Learning the parts of objects by nonnegative matrix factorization. Nature **401**, 788–791 (1999)
14. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. Proc. NIPS **13**, 556–562 (2001)
15. Cichocki, A., Zdunek, R., Phan, A.H., Amari, S.: Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation. Wiley, New York (2009)
16. Ozerov, A., Févotte, C.: Multichannel nonnegative matrix factorization in convolutive mixtures for audio source separation. IEEE Trans. ASLP **18**(3), 550–563 (2010)
17. Kameoka, H., Yoshioka, T., Hamamura, M., Roux, J., Kashino, K.: Statistical model of speech signals based on composite autoregressive system with application to blind source separation. In: Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R., Vincent, E. (eds.) LVA/ICA 2010. LNCS, vol. 6365, pp. 245–253. Springer, Heidelberg (2010). doi:10.1007/978-3-642-15995-4_31
18. Sawada, H., Kameoka, H., Araki, S., Ueda, N.: Multichannel extensions of nonnegative matrix factorization with complex-valued data. IEEE Trans. ASLP **21**(5), 971–982 (2013)
19. Murata, N., Ikeda, S., Ziehe, A.: An approach to blind source separation based on temporal structure of speech signals. Neurocomputing **41**(14), 1–24 (2001)
20. Ishimura, M., Makino, S., Yamada, T., Ono, N., Saruwatari, H.: Noise reduction using independent vector analysis and noise cancellation for a hose-shaped rescue robot. In: Proceedings of IWAENC (2016)
21. Hänsler, E., Schmidt, G.: Acoustic Echo and Noise Control: A Practical Approach. Wiley, New York (2004)
22. Vincent, E., Gribonval, R., Févotte, C.: Performance measurement in blind audio source separation. IEEE Trans. ASLP **14**, 1462–1469 (2006)