Human-Voice Enhancement based on Online RPCA for a Hose-shaped Rescue Robot with a Microphone Array

Yoshiaki Bando[1], Katsutoshi Itoyama[1], Masashi Konyo[2], Satoshi Tadokoro[2], Kazuhiro Nakadai[3], Kazuyoshi Yoshii[1], and Hiroshi G. Okuno[4]

***Abstract*— This paper presents an online real-time method that enhances human voices included in severely noisy audio signals captured by microphones of a hose-shaped rescue robot. To help a remote operator of such a robot pick up a weak voice of a human buried under rubble, it is crucial to suppress the loud ego-noise caused by the movements of the robot in real time. We tackle this task by using online robust principal component analysis (ORPCA) for decomposing the spectrogram of an observed noisy signal into the sum of low-rank and sparse spectrograms that are expected to correspond to periodic ego-noise and human voices. Using a microphone array distributed on the long body of a hose-shaped robot, ego-noise suppression can be further improved by combining the results of ORPCA applied to the observed signal captured by each microphone. Experiments using a 3-m hose-shaped rescue robot with eight microphones show that the proposed method improves the performance of conventional ego-noise suppression using only one microphone by 7.4 dB in SDR and 17.2 in SIR.**

I. Introduction

Hose-shaped rescue robots have been developed for gathering information in narrow spaces under collapsed buildings where humans or animals cannot go [1]–[3]. They have thin, long, and flexible bodies, and self-locomotion mechanisms. The Active Hose-II robot [1], for example, has small powered wheels enabling it to move forward, and the Active Scope Camera robot [2], [3] can move forward by vibrating cilia covering its body. In 2008 the Active Scope Camera robot was used in an actual search-and-rescue mission in Jacksonville, Florida, USA [4].

To avoid failing to hear the voice of a human at an unseen and distant place, real-time ego-noise suppression is crucial for a hose-shaped rescue robot. It should be noted that the rescue activity is a race against time. Although hose-shaped rescue robots are asked to keep moving for searching a wide area in a limited time, conventional robots need to stop their actuators and keep silent for hearing the voice of a human. This strategy, however, is inefficient and often misses a voice occurring when a robot is moving. Since the ego-noise of a robot changes dynamically according to the movements of wheels or vibrators and friction between the robot body and surrounding materials, it is difficult to use conventional ego-noise suppression methods [5]–[8] that assume the noise to be stable or learned in advance.

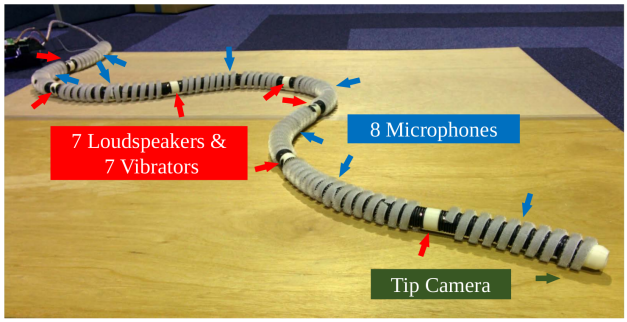


Fig. 1. Prototype hose-shaped rescue robot with eight-channel microphone array and driving mechanism.

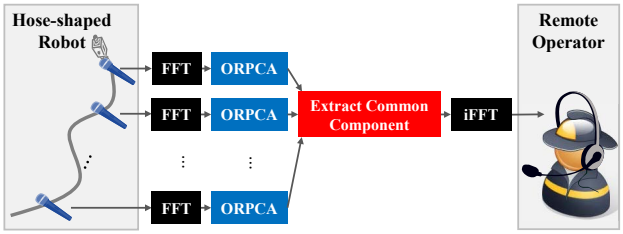


Fig. 2. Overview of proposed online ego-noise suppression system.

In a real rescue scenario, we need to deal with a severely-low signal-to-noise ratio (SNR) condition because ego-noise sounds generated from the body of a robot are much closer to microphones than human voices. One possibility is to use source separation methods based on microphone arrays [9], [10] for emphasizing human voices. These methods, which need the precise information of microphone positions, however, cannot be used because the shape of a hose-shaped robot changes flexibly in narrow gaps and it is difficult to know the positions of microphones on the robot [11], [12]. Although several methods independent of the microphone positions called blind source separation [13], [14] are proposed, these methods requires too much computational costs to use in real time.

A promising approach to unsupervised and efficient ego-noise suppression for a hose-shaped rescue robot is to use robust principal component analysis (RPCA) [15] that was originally proposed for decomposing an input matrix into the sum of sparse and low-lank matrices. Applying RPCA to an audio spectrogram, it can be decomposed into frequency components that appear repeatedly (*e.g.*, the ego-noise of a

[1]Graduate School of Informatics, Kyoto University, Kyoto, 606-8501, Japan {yoshiaki, itoyama, yoshii}@kuis.kyoto-u.ac.jp
[2]Graduate School of Information Science, Tohoku University, Miyagi, 980-8579, Japan {konyo, tadokoro}@rm.is.tohoku.ac.jp
[3]Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Saitama, 351-0114, Japan / Honda Research Institute Japan Co., Ltd. nakadai@jp.honda-ri.com
[4]Graduate Program for Embodiment Informatics, Waseda University, Tokyo 169-0072, Japan okuno@aoni.waseda.jp

Demo: http://winnie.kuis.kyoto-u.ac.jp/members/yoshiaki/demo/ssrr2015

TABLE I

STRONG AND WEAK POINTS OF PROPOSED METHOD AND RELATED WORK OF EGO-NOISE SUPPRESSION

Method	Non-stable noise	No prior learning	Independent of microphone positions
Ince et al. [5]			✓
Tezuka et al. [6]	✓		✓
Cauchi et al. [7]	✓		✓
Nakajima et al. [8]		✓	✓
Okutani et al. [18]	✓	✓	
Proposed	✓	✓	✓

hose-shaped rescue robot) and other components that occur infrequently (*e.g.*, the voice of a human) without prior learning [16]. RPCA can also suppress periodic ego-noise caused by wheels and vibrators because the noise spectrogram has a low-rank structure. Furthermore, periodic environmental noise (*e.g.*, noise derived from rescue helicopters and vehicles) are expected to be suppressed.

This paper presents a novel statistical method that can suppress the ego-noise of a hose-shaped rescue robot by applying an online version of RPCA (ORPCA) [17] to a microphone array distributed along the robot as shown in Fig. 1. Our method assumes that a target voice is similarly recorded by all microphones and the ego-noise is differently recorded by each microphone. This enables us to suppress the ego-noise in two steps: 1) suppress the ego-noise at each microphone by using ORPCA, and 2) extract the components common among the microphones by combining the results of ORPCA (Fig. 2). The effectiveness of the proposed method was evaluated using the prototype hose-shaped rescue robot with an eight-channel microphone array.

II. RELATED WORK

This section reviews related work on ego-noise suppression. We first introduce the ego-noise suppression methods for a single microphone and then review methods for a microphone array. The strong and weak points of each method are summarized in Table I.

A. Noise Reduction for Single Microphone

Ince et al. [5] developed an ego-noise suppression method for a single microphone using the sensor data of actuators. This method uses a noise database that stores pairs of motor-sensor values and corresponding ego-noise. The ego-noise is estimated by searching a noise template from the database with the sensor values. Although this method works well when the sensor data and ego-noise have a correlation, its performance is degraded by unknown noise sounds and unstable noise.

Tezuka et al. [6] and Cauchi et al. [7] estimated the ego-noise using extensions of non-negative matrix factorization (NMF). NMF decomposes the input matrix into basis and activation matrices [19]. The basis matrix is a set of template frequency components, and the activation matrix is the coefficients of the bases at each time frame. Since the NMF-based approach represents the noise spectrum with multiple bases, these methods can maintain non-stable ego-noise. This approach, however, needs to learn the basis of ego-noise in advance because it is difficult to distinguish the bases of ego-noise and a target voice.

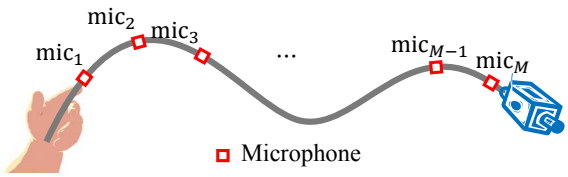


Fig. 3. Configuration of microphones on a hose-shaped rescue robot.

Nakajima et al. [8] developed a method for estimating the stable noise without prior learning. This method, called histogram-based recursive level estimation (HRLE), estimates the noise using the cumulative histogram of the input power spectrum. It therefore robustly estimates the stable noise even when there is some sudden and large-power signal. HRLE is implemented in an open source robot audition software called HARK [1] [9]. Since HRLE estimates the noise using only the histogram and ignores the error between the estimated noise and current input, it fails to estimate non-stable noise.

B. Noise Reduction for Microphone Array

Okutani et al. [18] developed an ego-noise-robust sound source localization method that uses a microphone array. It is called multiple signal classification based on incremental generalized eigenvalue decomposition (iGEVD-MUSIC). GEVD-MUSIC is an online sound source localization method suppressing the ego-noise given as a noise correlation matrix between the microphones [20]. iGEVD-MUSIC sequentially predicts the current noise correlation matrix from the previous input signal. Since the correlation matrix represents the relationship between microphones and noise source positions, this method is robust against the stationarity of the noise. Okutani et al. robustly estimated the position of the target sound source with a microphone array equipped on a multicopter even when it was flying. This method, however, assumes that the positions of the microphones are known in advance. There are also several methods that use a microphone array equipped a rescue robot, they need the microphone positions [21]–[23].

III. ONLINE EGO-NOISE SUPPRESSION

This section describes the proposed online ego-noise suppression method that extracts the components common among the microphones by combining the ORPCA results of the microphones at each frequency bin.

A. Problem Statement

We assume a hose-shaped rescue robot that has microphones distributed along its body as depicted in Fig. 3. The microphones are given an index ranging from 1 at the hand position to M at the tip of the robot. We define t as the time frame index, F as the number of frequency bins, and f as the frequency bin index. The other notations are summarized in Table II.

The ego-noise suppression problem for a hose-shaped robot is defined as follows:

[1]Honda Research Institute Japan Audition for Robots with Kyoto Univ.

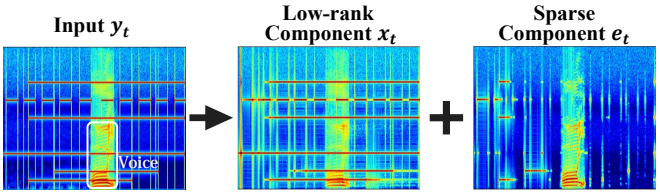


Fig. 4. Spectrograms of low-rank and sparse components extracted by ORPCA. Input is the mixture of a human voice and a simple noise (pure tones).

TABLE II

NOTATIONS OF MATHEMATICAL SYMBOLS

Symbol	Meaning
M	Number of microphones
F	Number of frequency bins
m	Microphone index ($m = 1, \cdots, M$)
f	Frequency bin index
t	Time frame index
k	The number of basis for low-rank component
$\boldsymbol{y}_{mt} \in \mathbb{R}^F$	Amplitude spectrum of input signal
$\boldsymbol{s}_t \in \mathbb{R}^F$	Amplitude spectrum of output signal
$\boldsymbol{x}_{mt} \in \mathbb{R}^F$	Low-rank component ($= \mathbf{L}_m \boldsymbol{r}_{mt}$)
$\boldsymbol{e}_{mt} \in \mathbb{R}^F$	Sparse component
$\boldsymbol{r}_{mt} \in \mathbb{R}^K$	Coefficients corresponding to the basis of low-rank component
$\mathbf{L}_m \in \mathbb{R}^{F \times K}$	Basis of low-rank components
$\boldsymbol{w}_{mt} \in \mathbb{R}^F$	Normalization coefficient.

Input: M-channel synchronized amplitude spectrums $\boldsymbol{y}_{1t}, \cdots, \boldsymbol{y}_{Mt} \in \mathbb{R}^F$
Output: Denoised amplitude spectrum $\boldsymbol{s}_t \in \mathbb{R}^F$

The input amplitude spectrum is defined as the absolute values of the short-time Fourier transform (STFT) of recorded signals. The time-domain signal of the output is obtained by conducting the inverse STFT to the product of the output amplitude spectrum $\boldsymbol{s}_t$ and the phase of the recorded signal.

B. Online RPCA

The proposed method uses an online extension of batch RPCA (ORPCA) [17]. The input signal of each channel $\boldsymbol{y}_{mt}$ is decomposed to low-rank component $\boldsymbol{x}_{mt}$ and sparse component $\boldsymbol{e}_{mt}$ by conducting ORPCA (Fig. 4):

$$\boldsymbol{y}_{mt} = \boldsymbol{x}_{mt} + \boldsymbol{e}_{mt}. \tag{1}$$

The ego-noise that periodically changes is assigned to the low-rank component, and the voice and other sparse noise are assigned to the sparse component [16].

To explain ORPCA which is independent of the microphone index m, in the rest of this section we leave out it. Let the $F \times t$ matrices of input spectra, low-rank components, and sparse components be $\mathbf{Y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_t]$, $\mathbf{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_t]$, and $\mathbf{E} = [\boldsymbol{e}_1, \cdots, \boldsymbol{e}_t]$. The original batch RPCA decomposes the input matrix into low-rank and sparse components by solving following problem [15]:

$$\min_{\mathbf{X},\mathbf{E}} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{X} - \mathbf{E}\|_F^2 + \lambda_1 \|\mathbf{X}\|_* + \lambda_2 \|\mathbf{E}\|_1 \right\} \tag{2}$$

where $\|\cdot\|_F$, $\|\cdot\|_*$, and $\|\cdot\|_1$ represent the Frobenius, nuclear, and L1 norms, respectively, and λ_1 and λ_2 represent the scale parameters. Since this optimization is solved by the augmented Lagrangian multiplier method and it performs SVD which accesses all samples of the input, it was difficult to solve the original RPCA problem in an online manner [15].

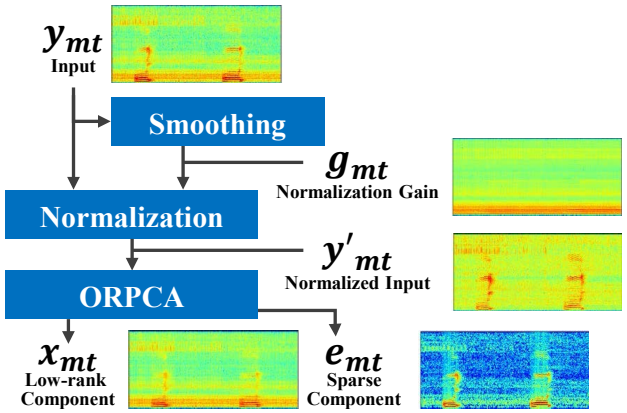


Fig. 5. Noise suppression using channel-wise ORPCA for multi-channel signals.

To solve the RPCA problem in an online manner, ORPCA solves the following alternative problem [17].

$$\min_{\mathbf{L},\mathbf{R},\mathbf{E}} \left\{ \frac{1}{2} \|\mathbf{Y} - \mathbf{L}\mathbf{R}^T - \mathbf{E}\|_F^2 + \frac{\lambda_1}{2} (\|\mathbf{L}\|_F^2 + \|\mathbf{R}\|_F^2) + \lambda_2 \|\mathbf{E}\|_1 \right\} \tag{3}$$

where $\mathbf{L} \in \mathbb{R}^{F \times K}$ and $\mathbf{R} \in \mathbb{R}^{t \times K}$ $(K < F, t)$ represent the bases of the low-rank components and the coefficients of the bases ($\mathbf{X} = \mathbf{L}\mathbf{R}^T$). As proved in [17], the local optima of this problem are the global optimal solutions of the original RPCA problem because the $\|\mathbf{X}\|_*$ is upper bounded by the $\mathbf{L}$ and $\mathbf{R}$ [24]:

$$\|\mathbf{X}\|_* = \|\mathbf{L}\mathbf{R}^T\|_* = \inf_{\mathbf{L},\mathbf{R}} \left\{ \frac{1}{2} \|\mathbf{L}\|_F + \frac{1}{2} \|\mathbf{R}\|_F \right\} \tag{4}$$

ORPCA solves the alternative RPCA problem (Eq. 3) by minimizing the following cost function obtained by transforming Eq. (3):

$$f_t(L) = \frac{1}{t} \sum_{t'=1}^{t} l(\boldsymbol{y}_{t'}, \mathbf{L}) + \frac{\lambda_1}{2t} \|\mathbf{L}\|_F^2 \tag{5}$$

$$l(\boldsymbol{y}_{t'}, \mathbf{L}) = \min_{\boldsymbol{r}_{t'}, \boldsymbol{e}_{t'}} \left\{ \frac{1}{2} \|\boldsymbol{y}_{t'} - \mathbf{L}\boldsymbol{r}_{t'} - \boldsymbol{e}_{t'}\|_2^2 + \frac{\lambda_1}{2} \|\boldsymbol{r}_{t'}\|_2^2 + \lambda_2 \|\boldsymbol{e}_{t'}\|_1 \right\} \tag{6}$$

where $l(\boldsymbol{y}_t, \mathbf{L})$ represents the loss function for each sample. The coefficients of the bases $\boldsymbol{r}_t$ and sparse component $\boldsymbol{e}_t$ are updated at each time frame by solving Eq. (6) with fixing the basis of low-rank subspace $\mathbf{L}$. The basis $\mathbf{L}$, on the other hand, is updated at each time frame by minimizing following objective function $g_t(\mathbf{L})$ with fixing $\boldsymbol{r}_t$ and $\boldsymbol{e}_t$:

$$g_t(\mathbf{L}) = \frac{1}{t} \sum_{t'=1}^{t} l'(\boldsymbol{y}_{t'}, \mathbf{L}) + \frac{\lambda_1}{2t} \|\mathbf{L}\|_F^2 \tag{7}$$

$$l'(\boldsymbol{y}_{t'}, \mathbf{L}) = \frac{1}{2} \|\boldsymbol{y}_{t'} - \mathbf{L}\boldsymbol{r}_{t'} - \boldsymbol{e}_{t'}\|_2^2 + \frac{\lambda_1}{2} \|\boldsymbol{r}_{t'}\|_2^2 + \lambda_2 \|\boldsymbol{e}\|_1 \tag{8}$$

This function provides an upper bound for the cost function $f_t(\mathbf{L})$. ORPCA solves Eqs. (6) and (7) using an off-the-shelf solver and block-coordinate descent with warm restarts, respectively [17].

C. Online Normalization of Input Signal

The ego-noise of our hose-shaped rescue robot has peaks at low frequency bins. Since ORPCA estimates the low-rank components with the same weight for the all frequency

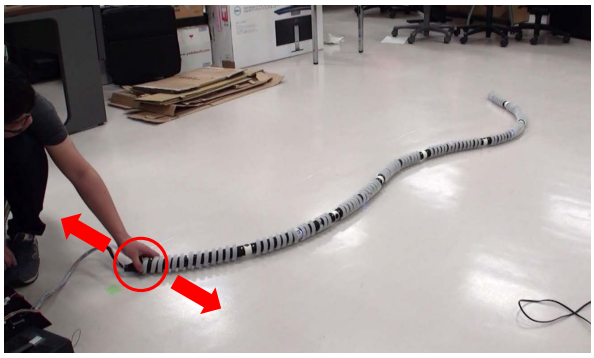


Fig. 6. Prototype hose-shaped robot placed on the floor of an office room.

bins (Eq. 7), it over-fits to the peaks. We therefore apply a normalization coefficient $\boldsymbol{w}_{mt} = [w_{mt1}, \cdots, w_{mtF}]^T \in \mathbb{R}^F$ to the input $\boldsymbol{y}_{mt}$:

$$y'_{mtf} = \frac{1}{w_{mtf}} y_{mtf}. \tag{9}$$

Since the peaks of the ego-noise changes depending on the environment around the robot, the proposed method learns the normalization coefficient in an online manner (Fig. 5). We assume that the average ego-noise does not change frequently and drastically, therefore the normalization coefficient is updated as follows:

$$\boldsymbol{w}_{mt} = \alpha \boldsymbol{y}_{mt} + (1-\alpha)\boldsymbol{w}_{m(t-1)} \tag{10}$$

where α is a learning weight parameter that is set to a small value (e.g., 1.0×10^{-2}).

D. Combining Online RPCA Results at Microphones

The sparse components of the microphones $\boldsymbol{e}_{mt}$ are integrated to extract the common component. Each sparse component has large sparse noise measured when the corresponding microphone touches on the environment, and small noise which is suppression residuals of the ego-noise. To be robust against these noise, the integration is conducted by taking a median at each frequency bin as follows:

$$s_{tf} = \mathrm{Median}(e_{1tf}, \cdots, e_{Mtf}) \quad \text{for all} \quad f = 1, \cdots, F \tag{11}$$

where $\mathrm{Median}(\cdots)$ represents a median of the arguments.

The whole proposed noise suppression algorithm is summarized as follows:

Algorithm 1 Proposed ego-noise suppression method

for $t = 1, 2, 3 \cdots$ **do**
  Observe the audio input signal $\boldsymbol{y}_{mt} (m = 1, \cdots, M)$
  for $m = 1, \cdots, M$ **do**
    Update normalizing coefficient $\boldsymbol{w}_{mt}$ (Eq. (10))
    Normalize input $\boldsymbol{y}_{mt}$ to get $\boldsymbol{y}'_{mt}$ (Eq. (9))
    Apply ORPCA to $\boldsymbol{y}'_{mt}$ to get sparse component $\boldsymbol{e}'_{mt}$
    Multiply $\boldsymbol{e}'_{mt}$ by the coefficient $\boldsymbol{w}_{mt}$ to get $\boldsymbol{e}_{mt}$
  end for
  for $f = 1, \cdots, F$ **do**
    Take a median of $[e_{1tf}, \cdots, e_{Mtf}]$ to get output s_{tf}
  end for
end for

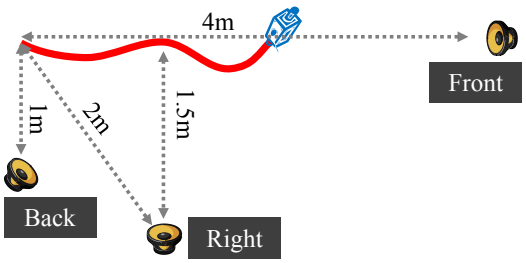


Fig. 7. Arrangement of a hose-shaped rescue robot and a loudspeaker. We tested three loudspeaker positions: front, right, and back.

IV. Experimental Evaluation

This section reports an experiment evaluating the proposed method of ego-noise suppression using a prototype hose-shaped rescue robot.

A. Experimental Conditions

Fig. 1 shows the prototype hose-shaped rescue robot used in this experiment. The body was made with a corrugated tube 38 mm in diameter and 3 m long. This robot had a self-propelling mechanism the same as that of the Active Scope Camera robot [3]. The entire surface of the robot was covered by cilia and the robot moved forward by vibrating the cilia. This robot had $M = 8$ microphones positioned on its body at a regular interval of 40 cm. The body was rotated 90 degrees after each of the microphones was installed with in order to avoid having all microphones obstructed by the ground.

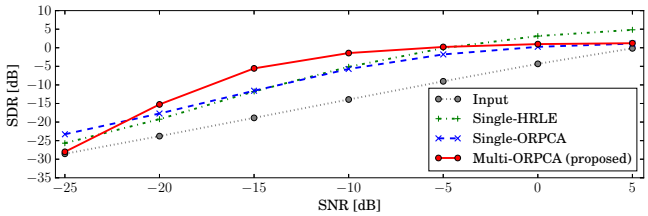
We recorded the ego-noise and target voice separately, and then mixed them with varying SNR from -25 dB to $+5$ dB. This experiment was conducted in an experimental room with a reverberation time (RT_{60}) of 740 ms (Fig. 6). The ego-noise was recorded for 60 seconds sliding the robot left and right by the vibrators and a hand. The arrangement of the robot and the loudspeaker from which the target voice emerged is shown in Fig. 7. We tested three positions of the loudspeaker: front, right, and back. A low-noise target voice signal was generated by convoluting the clean male voice recorded in an anechoic chamber and the impulse response recorded with the loudspeaker. These synchronized eight-channel acoustic signals were sampled at 16 kHz and 16 bits using HARK [9].

We compared following three methods:

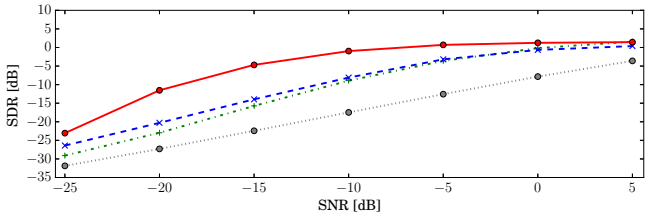
1) **Multi-ORPCA**: the proposed method using all the eight microphones.
2) **Single-ORPCA**: ORPCA using only the 8th (tip) microphone.
3) **Single-HRLE**: HRLE using only the 8th microphone.

HRLE is one of the conventional methods that solve the same noise suppression problem as ORPCA. The parameters of the proposed method were decided experimentally, and those of HRLE were set to the default values of the HARK implementation.

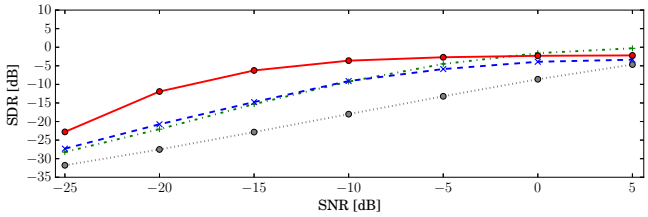
We implemented the proposed method as a HARK module by using C++ and a linear algebra library called Eigen3. The estimation was conducted with a standard desktop computer with an Intel Core i7–4790 CPU (4-core, 3.6 GHz) and 8.0 GB of memory. The elapsed time for the proposed



(a) Loudspeaker position: Front

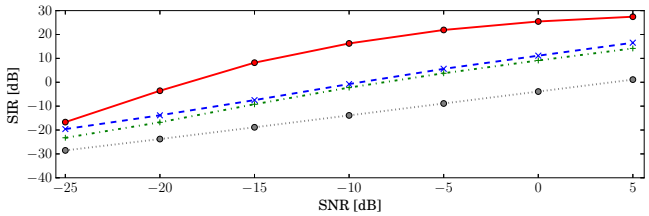


(b) Loudspeaker position: Right

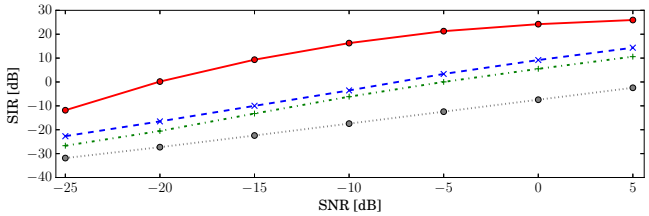


(c) Loudspeaker position: Back

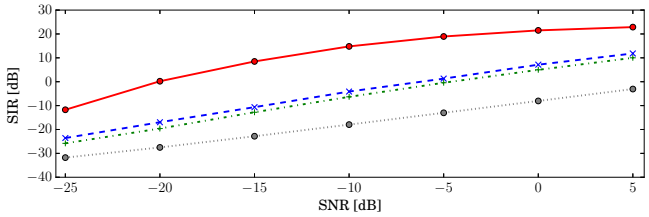
Fig. 8. SDRs of input audio signal at 8th (tip) microphone and signals denoised by Single-HRLE, Single-ORPCA, and Multi-ORPCA.



(a) Loudspeaker position: Front



(b) Loudspeaker position: Right



(c) Loudspeaker position: Back

Fig. 9. SIRs of input audio signal at 8th (tip) microphone and signals denoised by Single-HRLE, Single-ORPCA, and Multi-ORPCA.

method with an 8-ch input signal was 20.0 s. This value was small enough compared with the whole signal length of the input (60.0 s) that our method could work in real time.

The ego-noise suppression performance was evaluated using signal-to-distortion ratio (SDR) and signal-to-interference ratio (SIR) [25]. SDR measures the overall quality of the retrieved noise-suppressed signal, and SIR measures how much the interference due to ego-noise is suppressed. They were measured using a Python toolkit called MIR-EVAL [26].

B. Experimental Results

As shown in Fig. 8-(a), when the loudspeaker was in front of the robot, the SDR of the proposed method was higher than that of the other methods at the SNR conditions between -20 dB and -5 dB. Moreover, as shown in Figs. 8-(b) and 8-(c), when the loudspeaker was to the right of the robot or behind the robot, the SDR of the proposed method was higher than that of the other methods at all the SNR conditions -5 dB or less. The SIR of the proposed method was improved more than 2.9 dB compared with that of the other methods in the all conditions.

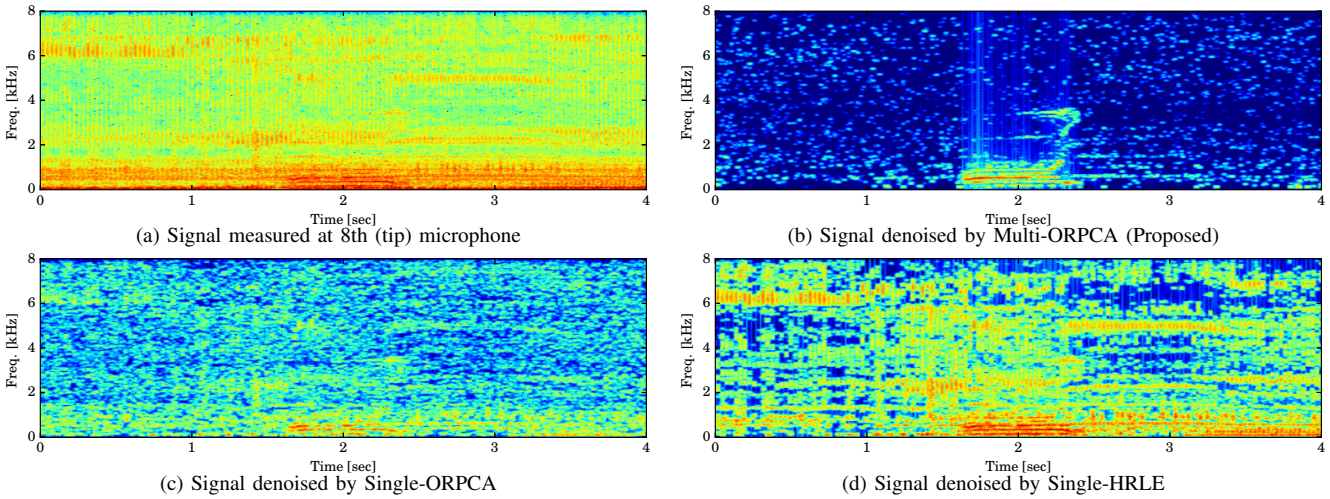
SIR measures how much the ego-noise is suppressed. As shown in Figs. 10 and 11, the suppressed spectrogram of the proposed method had lower noise than those of the others. The ego-noise changed with a 30-Hz frequency which was shown as vertical stripe patterns in Fig. 10-(a). The result of Single-HRLE remains the fluctuation residuals of the ego-noise as vertical stripe patterns (Fig. 10-(d)), whereas the result of the proposed method and Single-ORPCA (Figs. 10-(b) and (c)) were suppressed.

When the loudspeaker was in front of the robot, the SDRs and SIRs for the proposed method were not as good as they were when it was in the other two positions. This shows the suppression performance depends on the relative positions of the microphones and the target sound source. This is because the proposed method integrates all microphones with the same weights, whereas the power of the target sound was different at each microphone. One way to overcome this problem is to integrate the microphones by weighting them with the estimated SNR of the target voice.

V. Conclusion

This paper presented an online real-time method that suppresses the ego-noise of a hose-shaped rescue robot. The proposed method suppresses the ego-noise at each microphone by using ORPCA, and extracts the components common among the microphones by combining the ORPCA results. The experiments using a 3-m hose-shaped rescue robot with an eight-channel microphone array shows that the proposed method improves the performance of conventional ego-noise suppression using only one microphone by 7.4 dB in SDR and 17.2 in SIR. The results also show that the suppression performance is often degraded, depending on the relative positions of the microphones and the target sound source.

To improve the proposed method, we plan to develop a microphone array integration method that weights the microphones with the estimated SNR of the target voice. Since the proposed method suppresses the noise that frequently appears, it can suppress not only the ego-noise of a hose-shaped rescue robot but also other environmental noise and the ego-



(a) Signal measured at 8th (tip) microphone

(b) Signal denoised by Multi-ORPCA (Proposed)

(c) Signal denoised by Single-ORPCA

(d) Signal denoised by Single-HRLE

Fig. 10. Examples of spectrograms obtained when the loudspeaker is to the right of the robot and SNR is set to -5 dB.

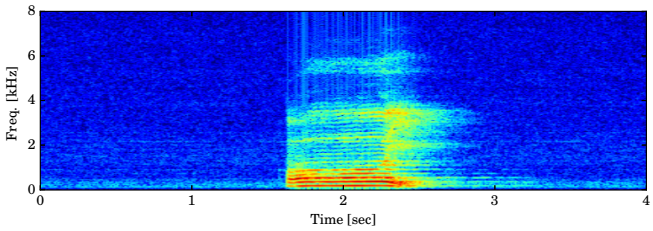


Fig. 11. Spectrogram of target voice for the condition of Fig. 10

noise of other noisy rescue robots without prior learning. Our future work includes investigating the method's performance with other rescue robots, such as multicopters [18]. Furthermore, to evaluate the effectiveness in the search-and-rescue tasks, we will conduct more practical experiments in simulated disaster sites.

ACKNOWLEDGMENTS

This study was partially supported by ImPACT Tough Robotics Challenge and by JSPS KAKENHI No. 24220006 and No. 15J08765.

REFERENCES

[1] A. Kitagawa *et al.*, "Development of small diameter active hose-ii for search and life-prolongation of victims under debris," *Journal of Robotics and Mechatronics*, vol. 15, no. 5, pp. 474–481, 2003.

[2] H. Namari *et al.*, "Tube-type Active Scope Camera with high mobility and practical functionality," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2012, pp. 3679–3686.

[3] J. Fukuda *et al.*, "Remote vertical exploration by Active Scope Camera into collapsed buildings," in *IEEE/RSJ IROS*, 2014, pp. 1882–1888.

[4] S. Tadokoro *et al.*, "Application of active scope camera to forensic investigation of construction accident," in *IEEE International Workshop on Advanced Robotics and its Social Impacts*, 2009, pp. 47–50.

[5] G. Ince *et al.*, "Incremental learning for ego noise estimation of a robot," in *IEEE/RSJ IROS*, 2011, pp. 131–136.

[6] T. Tezuka *et al.*, "Ego-motion noise suppression for robots based on semi-blind infite non-negative matrix factorization," in *IEEE International Conference on Robotics and Automation*, 2014, pp. 6293–6298.

[7] B. Cauchi *et al.*, "Reduction of non-stationary noise for a robotic living assistant using sparse non-negative matrix factorization," in *Workshop on Speech and Multimodal Interaction in Assistive Environments*, 2012, pp. 28–33.

[8] H. Nakajima *et al.*, "An easily-configurable robot audition system using histogram-based recursive level estimation," in *IEEE/RSJ IROS*, 2010, pp. 958–963.

[9] K. Nakadai *et al.*, "Design and implementation of robot audition system HARK – open source software for listening to three simultaneous speakers," *Advanced Robotics*, vol. 24, no. 5-6, pp. 739–761, 2010.

[10] V. M. Valin *et al.*, "Enhanced robot audition based on microphone array source separation with post-filter," in *IEEE/RSJ IROS*, 2004, pp. 2123–2128.

[11] M. Ishikura *et al.*, "Shape estimation of flexible cable," in *IEEE/RSJ IROS*, 2012, pp. 2539–2546.

[12] Y. Bando *et al.*, "A sound-based online method for estimating the time-varying posture of a hose-shaped robot," in *IEEE International Symposium on Safety, Security, and Rescue Robotics*, 2014, pp. 1–6.

[13] J. Taghia *et al.*, "A variational Bayes approach to the underdetermined blind source separation with automatic determination of the number of sources," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012, pp. 253–256.

[14] H. Sawada *et al.*, "Underdetermined convolutive blind source separation via frequency bin-wise clustering and permutation alignment," *IEEE Transaction on Audio, Speech and Language Processing (TASLP)*, vol. 19, no. 3, pp. 516–527, 2011.

[15] E. J. Candès *et al.*, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, p. 11, 2011.

[16] C. Sun *et al.*, "Noise reduction based on robust principal component analysis," *Journal of Computational Information Systems*, vol. 10, no. 10, pp. 4403–4410, 2014.

[17] J. Feng *et al.*, "Online robust PCA via stochastic optimization," in *Annual Conference on Neural Information Processing Systems*, 2013, pp. 404–412.

[18] K. Okutani *et al.*, "Outdoor auditory scene analysis using a moving microphone array embedded in a quadrocopter," in *IEEE/RSJ IROS*, 2012, pp. 3288–3293.

[19] S. A. Abdallah *et al.*, "Polyphonic music transcription by non-negative sparse coding of power spectra," in *International Conference on Music Information Retrieval (ISMIR)*, 2004, pp. 318–325.

[20] K. Nakamura *et al.*, "Intelligent sound source localization for dynamic environments," in *IEEE/RSJ IROS*, 2009, pp. 664–669.

[21] M. Basiri *et al.*, "Robust acoustic source localization of emergency signals from micro air vehicles," in *IEEE/RSJ IROS*, 2012, pp. 4737–4742.

[22] H. Sun *et al.*, "A far field sound source localization system for rescue robot," in *IEEE International Conference on Control, Automation and Systems Engineering*, 2011, pp. 1–4.

[23] J. C. Murray *et al.*, "Robotic sound-source localisation architecture using cross-correlation and recurrent neural networks," *Neural Networks*, vol. 22, no. 2, pp. 173–189, 2009.

[24] B. Recht *et al.*, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[25] E. Vincent *et al.*, "Performance measurement in blind audio source separation," *IEEE TASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.

[26] C. Raffel *et al.*, "mir_eval: A transparent implementation of common MIR metrics," in *ISMIR*, 2014, pp. 367–372.