*ICIC Express Letters*
Volume **14**, Number **4**, April **2020**

ICIC International ©2020 ISSN 1881-803X
pp. **337–345**

# A BASIC STUDY OF NOISE REDUCTION ON THE ANALYSIS OF BURST GRAVITATIONAL WAVES BY DIRECT AND PARALLEL DENOISING AUTOENCODER

Hiroyuki Hayashi[1], Kazuki Sakai[2], Hiroyuki Hamazumi[1]
Hirotaka Takahashi[3] and Yuto Omae[4]

[1]Department of Electrical and Electronic Engineering
National Institute of Technology, Tokyo College
1220-2 Kunugida, Hachioji, Tokyo 193-0997, Japan
{ s18707; hamazumi }@tokyo.kosen-ac.jp

[2]Department of Electronic Control Engineering
National Institute of Technology, Nagaoka College
888 Nishikatakai, Nagaoka, Niigata 940-8532, Japan
k-sakai@nagaoka-ct.ac.jp

[3]Department of Information and Management System Engineering
Nagaoka University of Technology
1603-1 Kamitomioka, Nagaoka, Niigata 940-2188, Japan
hirotaka@kjs.nagaokaut.ac.jp

[4]Department of Industrial Engineering and Management
College of Industrial Technology
Nihon University
1-2-1 Izumi, Narashino, Chiba 275-8575, Japan
oomae.yuuto@nihon-u.ac.jp

Received October 2019; accepted January 2020

Abstract. *Reconstruction of a burst gravitational waveform by denoising observed noisy data is one of the essential issues of gravitational wave astronomy. Conventional denoising methods require the knowledge of the frequency bands of the noise which is contained in observed data, but it is difficult to understand the statistical properties of the noise of observed data because it is known that the noise has non-Gaussian and non-stationary properties. In this paper, we propose direct and parallel denoising autoencoder for high-quality denoising without such kind of knowledge, and demonstrate our algorithm to reconstruct one of the typical models of burst gravitational waveforms from the noisy data.*
**Keywords:** Machine learning, Denoising autoencoder, Gravitational wave

1. **Introduction.** Gravitational waves (GWs) are ripples of spacetime and propagate space at the same speed as lights like electromagnetic waves propagate space as oscillations of an electromagnetic field. Whereas electromagnetic waves are emitted by accelerated electrically-charged particles, GWs are emitted by accelerated massive objects. Since the interaction between GWs and matters is weak, GWs can pass through almost all substances. It means that GWs have a potential to provide information of the inside of highly dense matters which other cosmic rays, such as electromagnetic waves and neutrinos, cannot come out.

Until now, the laser interferometer gravitational wave detectors, LIGO in the US and Virgo in Europe, have succeeded to observe 12 GWs. Those GWs were originated by

DOI: 10.24507/icicel.14.04.337

mergers of binary black holes and binary neutron stars [1]. Large-scale Cryogenic Gravitational Wave Telescope in Japan, which is named KAGRA [2], is being prepared to join the international GW observation network with its full configuration in early 2020.

The main targets of those detectors are the following three kinds of GWs: GWs from mergers of compact star binaries, continuous GWs, and burst GWs. Since the waveforms can be theoretically simulated for the first two kinds of GWs, we can use the matched filtering technique [3] in the data analysis. However, the precise simulation of the waveforms of burst GWs has not been completed yet. Therefore, the matched filtering technique cannot be applied to the analysis of them. A typical source of burst GWs is a supernova explosion. The observation of a GW from a supernova will give us valuable information about the dynamics of the explosion.

The data analysis of GWs consists of two steps, detection and parameter estimation. In the detection step, we search for the candidate of gravitational waves from enormous observation data containing various noises, such as the laser quantum noise, seismic noise from the ground. The various detection methods have been proposed depending on whether the waveforms are predictable or not. They are summarized in [3]. If we find a candidate of GWs in the detection step, we move on to the second step. In the parameter estimation step, we perform more detail analysis on the time segment containing the detection candidate. In particular, for the analysis of burst GWs, it is important to reconstruct signal waveforms by denoising from the noisy observed data. If we can reconstruct the signal waveform accurately, we can extract valuable information from the waveform and give it to researchers on both theory and simulation of supernova explosions.

Conventional denoising methods require the knowledge of the frequency bands of the noise which is contained in observed data to design filters. However, it is difficult to understand the statistical properties of the noise of observed data from GW detectors because it is known that the noise has non-Gaussian and non-stationary properties. On the other hand, in the field of deep learning, Vincent et al. [4] proposed the denoising autoencoder (DAE). Unlike conventional methods, DAE does not require such kind of knowledge. DAE has been applied to various fields: the reverberation suppression of sound source in voice recognition [5], denoising of the electrocardiogram waveform [6], the image recognition processing on moving object tracking system [7], and so on. DAE has become an important technology at a scene which requires denoising.

We propose a new model for high-quality denoising named direct and parallel DAE. We develop the model with the aim to apply it to the reconstruction of a burst gravitational

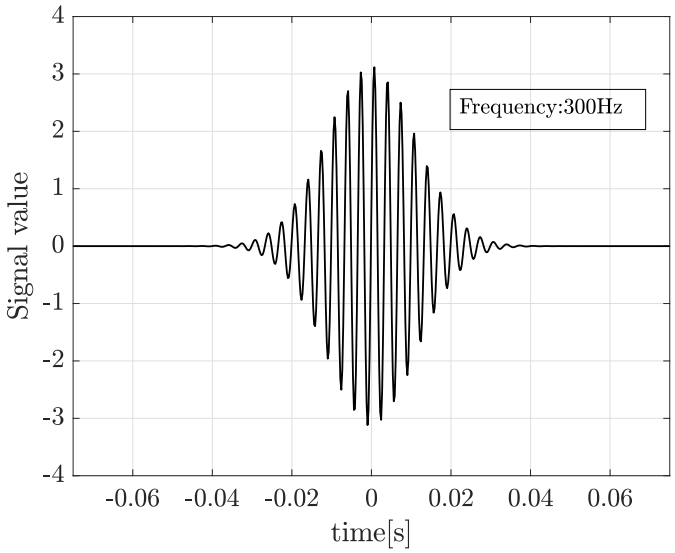


FIGURE 1. Sine-Gaussian signal

waveform from obtained noisy data. In this paper, as the first step, we focus on a sine-Gaussian waveform as the target waveform shown in Figure 1, which is one of the typically used simple models of burst GWs. We demonstrate our proposed model's ability to perform high-quality denoising and the reconstruction of the target waveform from noisy data with considerable accuracy by a simulation.

The paper is organized as follows. In Section 2, we explain our proposed algorithm of the direct and parallel DAE. In Section 3, we overview the simulation to verify the reliability of our method, and also its results and discussion are presented. Section 4 is devoted to a summary.

## 2. **Denoising Autoencoder.**

2.1. **Denoising process.** This section describes a denoising process by a neural network and its learning method. $\boldsymbol{s}$ is an original signal whose length is $n_1$. $\boldsymbol{x}$ is a noise-added signal, which is the superposition of $\boldsymbol{s}$ and noise. We consider constructing DAE which estimates $\boldsymbol{s}$ as $\hat{\boldsymbol{s}}$ after inputting $\boldsymbol{x}$ into it. Now, we make a DAE by a three-layered neural network shown in Figure 2. The units of the input, hidden, and output layers are denoted as $\boldsymbol{x}^1$, $\boldsymbol{x}^2$, and $\boldsymbol{x}^3$, whose numbers of dimensions are $n_1$, $n_2$, and $n_1$, respectively. Here, since $\boldsymbol{x}$ is the input and $\hat{\boldsymbol{s}}$ is the output of the network, $\boldsymbol{x}^1, \boldsymbol{x}^3$ are corresponding to $\boldsymbol{x}, \hat{\boldsymbol{s}}$, respectively. For reducing noise from input data, the dimensions of the input layer and hidden layer have to fulfill the condition of $n_1 > n_2$.

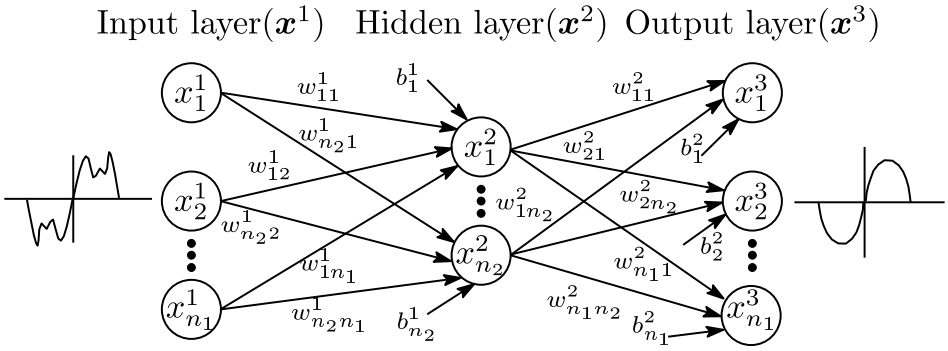


FIGURE 2. Three-layered neural network

Next, we describe how the noise reduction is achieved by the network. The transformation from $\boldsymbol{x}^1$ to $\boldsymbol{x}^2$ and from $\boldsymbol{x}^2$ to $\boldsymbol{x}^3$ are expressed by

$$\boldsymbol{x}^2 = \varsigma\left(\boldsymbol{W}^1\boldsymbol{x}^1 + \boldsymbol{b}^1\right), \tag{1}$$

$$\boldsymbol{x}^3 = \varsigma\left(\boldsymbol{W}^2\boldsymbol{x}^2 + \boldsymbol{b}^2\right), \tag{2}$$

where $\boldsymbol{W}^1$, $\boldsymbol{W}^2$, $\boldsymbol{b}^1$, and $\boldsymbol{b}^2$ are the model parameters, which are $n_2 \times n_1$ matrices, $n_1 \times n_2$ matrices, $n_2$ vectors, and $n_1$ vectors, respectively, and $\varsigma(z)$ is the sigmoid function. The elements of the model parameters are corresponding to $w_{jk}^i$ and $b_j^i$ in Figure 2.

2.2. **Learning method.** To make the network learn, we prepare $N_{\text{train}}$ realizations of noise $\boldsymbol{n}^{(i)}$ $(i = 1, \ldots, N_{\text{train}})$ and noise-added data $\boldsymbol{x}^{(i)} = \boldsymbol{s} + \boldsymbol{n}^{(i)}$ as training data, and define the cost function $L$ for parameter learning. First, we define $L^{(i)}$ as an evaluation value of $\hat{\boldsymbol{s}}^{(i)}$, the reconstructed waveform from $i$th noise-added data, which is expressed by,

$$L^{(i)} = \frac{1}{n_1}\sum_{m=1}^{n_1}\left(\hat{s}_m^{(i)} - s_m\right)^2, \tag{3}$$

where $\hat{s}_m^{(i)}$ and $s_m$ are the $m$-th values of $\hat{\boldsymbol{s}}^{(i)}$ and $\boldsymbol{s}$, respectively. In other words, Equation (3) is a mean squared error of the estimation of $\hat{\boldsymbol{s}}^{(i)}$. Then, the cost function $L$ is defined by

$$L = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} L^{(i)}. \tag{4}$$

It means the cost function is defined as the average of the evaluation values. The parameters of the network should be determined to minimize $L$. We update model parameters to be optimized by using the back-propagation [8].

2.3. **Direct and parallel DAEs.** We propose direct and parallel DAE to obtain higher-quality denoising than those by a single DAE. To construct our proposed model, we make four different DAEs which are named DAE1, DAE2, DAE3, and DAE4. Then, we construct three models as shown in Figure 3. (A) D-type1 is a model of connecting DAE1 and DAE2 in direct. (B) D-type2 is a model of connecting DAE3 and DAE4 in direct. (C) DP-type is a model of connecting D-type1 and D-type2 in parallel. In the (C) in Figure 3, the juncture point means that it outputs the average of input signals.

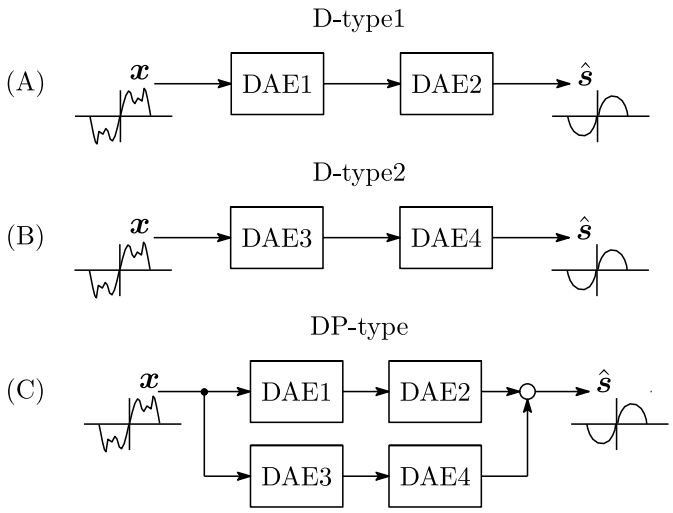


FIGURE 3. Models of the proposed method

Connecting DAEs in direct has the same effect as increasing hidden layers of a neural network, that is, it enables more complicated denoising processing. Connecting DAEs in parallel has the same effect as introducing the ensemble learning. In other words, the DAEs compensate for their estimation errors each other. For the reasons, we hypothesized that the DP-type, which combines the both advantages, performs higher-quality denoising than other types including each single DAE. In the next section, we evaluate the performance of each type by a simulation.

## 3. Simulation.

3.1. **Simulation outline.** First, we prepared training dataset and test dataset. The training dataset is for parameter learning and the test dataset is for model evaluation. To prepare the datasets, original signals and some realizations of noise are needed. As original signals, we use a sine-Gaussian signal $\boldsymbol{s}_1$ whose central frequency $f = 300$ Hz and a bias signal $\boldsymbol{s}_2$. The bias signal is used to evaluate the performance to noise-only data. They are expressed by,

$$s_{1,m} = 0.468\mathrm{e}^{-(62.5(m-m_0)\Delta t)^2}\sin(2\pi f(m-m_0)\Delta t) + 0.5, \tag{5}$$

$$s_{2,m} = 0.5, \tag{6}$$

where $m_0 = n_1/2$, and $\Delta t = 1/4096$ s is the sampling interval. The signal waveforms of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ are shown in Figure 4. As noises to be added to the original signals, we prepare 220 realizations of Gaussian noise whose standard deviation and mean value are 1 and 0, respectively. Each realization of noise is added to $\boldsymbol{s}_1$ or $\boldsymbol{s}_2$ to generate noise-added datasets $\boldsymbol{x}_1^{(i)}$ and $\boldsymbol{x}_2^{(j)}$ $(i, j = 1, \ldots, 110)$. We assigned each 100 data to the training dataset and each 10 data to the test dataset.

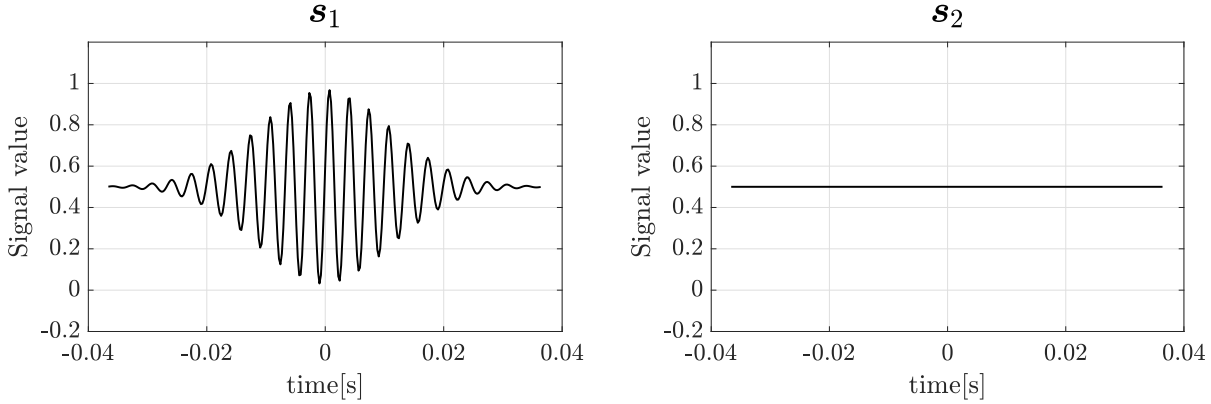


FIGURE 4. Signal waveforms of $\boldsymbol{s}_1$, $\boldsymbol{s}_2$

Next, we carry out parameter learning by using the training dataset. In this research, the dimensions of each DAE are set as $n_1 = 300$, $n_2 = 200$, and the iteration number of updating is set to be $10^4$. To remove noise components contained in the input data and extract the essence of the signal, the dimension of the hidden layer should be smaller than that of the input layer. However, if the dimension of the hidden layer is too small, the essence of the signal would be also removed. Based on the viewpoints, we set $n_2 = n_1 \times 2/3 = 200$ in this paper. We also initialize the model parameters of DAEs (DAE1 – DAE4) by random numbers following Gaussian distribution whose standard deviation and mean value are 0.5 and 0, respectively. After that, we evaluate each model by using the test dataset. To evaluate the denoising performances, we defined the degree of noise reduction $E$ [dB]. First, we calculate $P_{0,k}^{(i)}$ and $P_k^{(i)}$, $k = \{1, 2\}$, which are the power of the $i$th noise added to $\boldsymbol{s}_k$ and the power of the remaining noise contained in $i$th denoised data $\hat{\boldsymbol{s}}_k$. They are defined by

$$P_{0,k}^{(i)} = \sum_{m=1}^{n_1} \left( x_{k,m}^{(i)} - s_{k,m} \right)^2 = \sum_{m=1}^{n_1} \left( n_{k,m}^{(i)} \right)^2, \tag{7}$$

$$P_k^{(i)} = \sum_{m=1}^{n_1} \left( \hat{s}_{k,m}^{(i)} - s_{k,m} \right)^2. \tag{8}$$

Then, we calculate the average of them as $P_{0,k}$ and $P_k$, which are expressed by

$$P_{0,k} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} P_{0,k}^{(i)}, \tag{9}$$

$$P_k = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} P_k^{(i)}, \tag{10}$$

where $N_{\text{test}}$ is the number of test data, which is 10 in this research. Finally, we calculate the degree of noise reduction for each of $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$ defined by

$$E_k = 20 \log_{10} \frac{P_k}{P_{0,k}}. \tag{11}$$

As the evaluation, we use the average value of the degree of noise reduction for both waveforms,

$$E = \frac{E_1 + E_2}{2}. \tag{12}$$

If $E$ takes a negative value, it means the power of noise decreased after the denoising process. The smaller the value of $E$ is, the higher-quality denoising the model has achieved.

3.2. **Results and discussion.** In this section, we report the results of the noise reduction simulation. Figure 5 shows the change of the evaluation function $L$ along with updates of the learning process of DAE1. We can confirm that $L$ successfully converged, so that proper model parameters were obtained after the iteration.

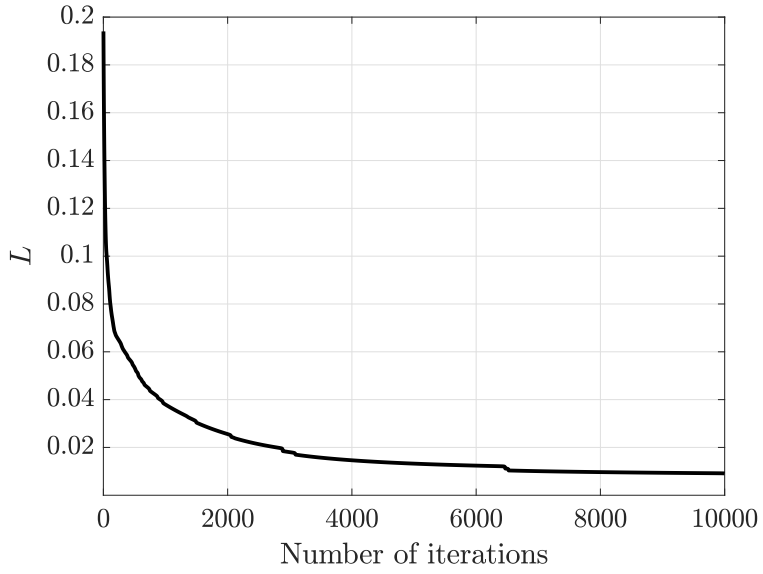


FIGURE 5. Change of $L$ of DAE1 at its learning

As the visualization of the noise reduction results, $\hat{\boldsymbol{s}}_1^{(i)}$ and $\hat{\boldsymbol{s}}_2^{(j)}$ by each model for specific $i$, $j$ are shown in Figure 6 with $\boldsymbol{x}_1^{(i)}$, $\boldsymbol{x}_2^{(j)}$. Focusing on the reconstructed waveforms by DAE1 and DAE2, it seems that not so much noise has been reduced. On the other hand, we can confirm that noise has been reduced to a certain extent by DAE3 and DAE4. Compared to these single DAEs, proposed models (D-type and DP-type) perform higher-quality denoising. In particular, we can see that DP-type succeeded to reconstruct the original waveforms with considerable accuracy.

However, even though DP-type could denoise with considerable accuracy, the noise has not been completely removed. To improve denoising performance, it is necessary to increase training data. Or, to increase the number of DAEs composing DP-type may also improve the performance, such as $3 \times 3$ DP-type or $4 \times 4$ DP-type or so. In addition, spike-like signals that do not exist in the input data appear in the output of DAEs. It can be caused by huge values in the model parameters $\boldsymbol{W}$ or $\boldsymbol{b}$, which are obtained due to overfitting to the training data, since the output layer uses the sigmoid function as its activation. One of the main future tasks is to prevent this problem by using the identity function instead of the sigmoid function at the output layer and/or introducing the L2 regularization in the parameter learning.

The evaluation value $E$ of each of the models, DAE1 – DAE4, D-type1, D-type2, and DP-type, are listed in Table 1. By comparing DAE1 – DAE4, it can be seen that the denoising performance differs from each other. $E$ is the largest for DAE1 and the smallest for DAE4. It means that the choice of initial values of model parameters has a certain influence on the denoising performance. Another fact to be noted is that $E$ of D-type1 is

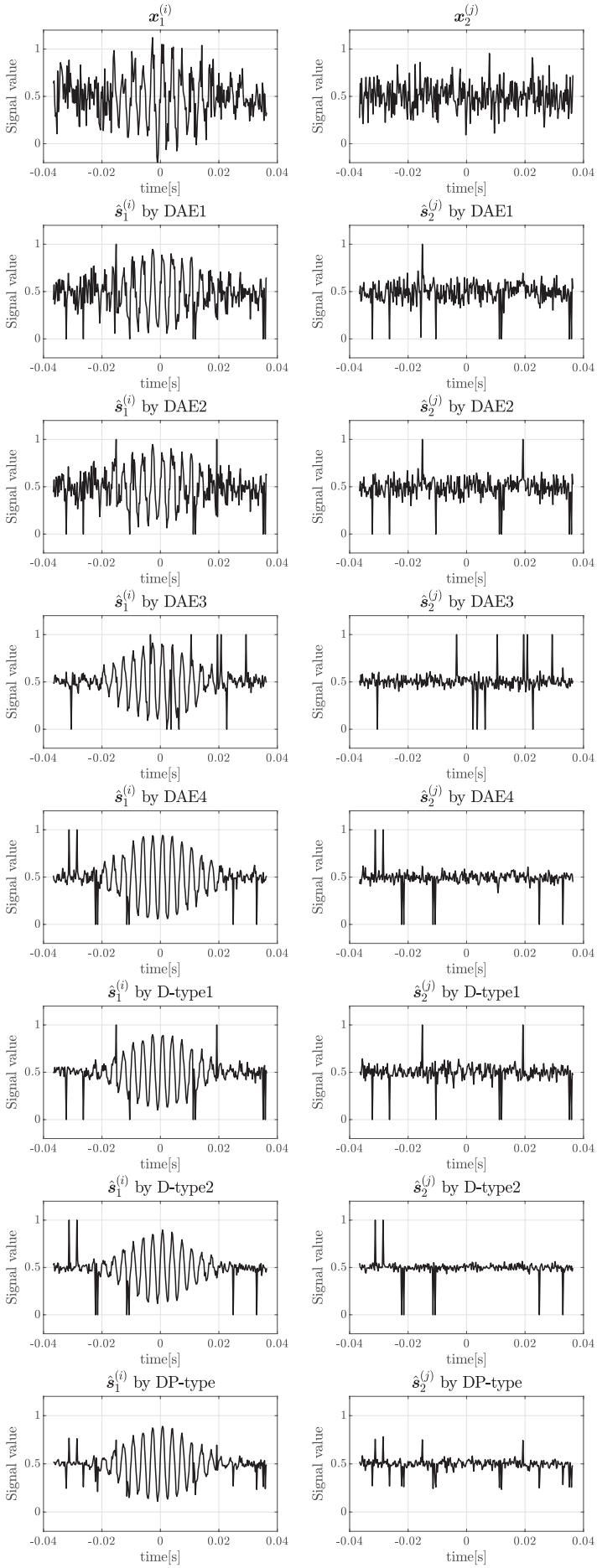


FIGURE 6. $\boldsymbol{x}_1^{(i)}$, $\boldsymbol{x}_2^{(j)}$ and $\hat{\boldsymbol{s}}_1^{(i)}$, $\hat{\boldsymbol{s}}_2^{(j)}$ by each model for specific $i$, $j$

TABLE 1. $E$ of each model

| Model | Structure | $E$ [dB] |
|---|---|---|
| DAE1 | Figure 2 | −2.669 |
| DAE2 | Figure 2 | −2.746 |
| DAE3 | Figure 2 | −7.340 |
| DAE4 | Figure 2 | −8.681 |
| D-type1 | Figure 3(A) | −7.387 |
| D-type2 | Figure 3(B) | −9.821 |
| DP-type | Figure 3(C) | −14.31 |

greater than the sum of $E$ of DAE1 and DAE2. This fact suggests that the direct connection itself has the potential to improve performance. Furthermore, DP-type recorded the smallest evaluation value of the proposed models. Consequently, we confirmed that making direct and parallel connections of DAE improves denoising performance.

4. **Conclusion.** We proposed a new denoising model based on the denoising autoencoder (DAE) for high-quality denoising with the aim to apply it to the data analysis of burst gravitational waves. Our proposed model, which is called direct and parallel DAE, is constructed by connecting DAEs in direct and parallel. For the evaluation of our proposed model, we conducted a noise reduction simulation with Gaussian noise and sine-Gaussian signal, which is typically used as a simple model of a burst gravitational wave. In the simulation, our proposed model performed highest-quality denoising and it succeeded to reconstruct the original waveforms with considerable accuracy. As the results, we found the facts that the choice of initial values of model parameters have a certain influence on the performance and making direct and parallel connections of DAE improves denoising performance. However, in this research, even the direct and parallel DAE is not able to remove noise, or reconstruct original signal, completely.

As the future works to improve the performance, we are planning to increase the training data and the iteration number of parameter updates at the learning process. In addition, we plan to use the identity function at the output layer and/or introducing the L2 regularization in the parameter learning. After that, we are going to investigate the effects of the increase in the number of DAEs composing a direct and parallel DAE. In this paper, we used only two kinds of signals for the simulation. For the next step, we will add more kinds of signals, such as sine-Gaussian signals with different central frequencies or real simulated gravitational waves from supernova explosions, in the training and test datasets. Furthermore, the actual noise from gravitational wave detectors is known to have non-Gaussian and non-stationary properties. For this reason, we would like to proceed with the performance evaluation with actual noise data obtained from gravitational wave detectors. The current model is presumed not to be effective for non-stationary noise because all the noises used in the parameter learning follow the same distribution. We consider that the effectiveness against non-stationary noise can be improved by adding noises which follow various distributions to the training data.

**Acknowledgment.** This work was supported in part by JSPS Grant-in-Aid for Young Scientists (B) (Grant No. 17K13179; Y. Omae), Young Scientists (Grant No. 19K14717; K. Sakai, and Grant No. 19K20062; Y. Omae), JSPS Grant-in-Aid for Scientific Research (C) (Grant No. 17K05437; H. Takahashi) and JSPS KAKENHI (Grant No. 17H06358; H. Takahashi).

## REFERENCES

[1] The LIGO Scientific Collaboration and the Virgo Collaboration, GWTC-1: A gravitational-wave transient catalog of compact binary mergers observed by LIGO and Virgo during the first and second observing runs, *Physical Review X*, vol.9, pp.031040-1-031040-49, 2019.

[2] The KAGRA Collaboration, KAGRA: 2.5 generation interferometric gravitational wave detector, *Nature Astronomy*, vol.3, pp.35-40, 2019.

[3] H. Shinkai, Direct detection and data analysis of gravitational waves, *Systems, Control and Information*, pp.370-375, 2018 (in Japanese).

[4] P. Vincent, H. Larochelle, Y. Bengio and P. Manzagol, Extracting and composing robust features with denoising autoencoders, *Proc. of the 25th International Conference on Machine Learning*, pp.1096-1103, 2008.

[5] T. Komiyama, N. Ishii, T. Shinozaki, Y. Horiuchi and S. Ishiguro, Examination of large vocabulary voices recognition under reverberant using denoising autoencoder, *IPSJ SIG Technical Report*, vol.2013-SLP-97, no.1, pp.1-6, 2013 (in Japanese).

[6] S. Miyatani, H. Koichi and M. Kano, Denoising autoencoder-based modification method for RRI data with premature ventricular contraction, *The 32nd Annual Conference of the Japanese Society for Artificial Intelligence*, 2J3-04, 2018 (in Japanese).

[7] N. Wang and D. Yeung, Learning a deep compact image representation for visual tracking, *Advances in National Information Processing Systems*, vol.26, pp.809-817, 2013.

[8] H. Hayashi and Y. Omae, On a direct and parallel denoising autoencoder and its evaluation, *Proc. of the 34th Fuzzy System Symposium*, pp.712-717, 2018 (in Japanese).