Neural-network-based microphone-array system trained with temporal-spatial patterns of multiple sinusoidal signals

Akihiro Iseki*, Kenji Ozawa† and Yuichiro Kinoshita‡

Interdisciplinary Graduate School, University of Yamanashi, 4–3–11 Takeda, Kofu, 400–8511 Japan

(Received 29 February 2016, Accepted for publication 23 August 2016)

Abstract: Although a previous study indicated that a microphone-array system consisting of seven microphones and a neural network realizes sharp sensitivity, it is only effective for a single frequency. In this work, we propose a new system with a modified input structure. Unlike the previous system, which was trained with the spatial patterns for a single frequency, the proposed system is trained with temporal-spatial patterns of the sound pressure distributions for sinusoidal signals at multiple frequencies. Three frequencies (425, 850, and 1,700 Hz) are used for the training process of a neural network in the proposed system. A computational simulation shows that the proposed system can realize sharp sensitivity with a half width of 5° at 425–1,700 Hz including untrained frequencies. Moreover, in an examination using an amplitude-modulated (AM) or frequency-modulated (FM) wave as the input signal, the proposed system achieves a higher performance than those in the previous study.

Keywords: Microphone array, Neural network, Sharp sensitivity, Temporal-spatial pattern

PACS number: 43.60.Fg, 43.60.Np [doi:10.1250/ast.38.63]

1. INTRODUCTION

A sharp directive microphone can effectively suppress noises in various situations such as recording and measuring a target sound in a noisy environment and sound source localization. A sharp directive microphone is usually constructed as a microphone array. Examples include delay-and-sum arrays [1,2] and adaptive microphone arrays [3,4]. Such linear systems require a number of microphones and have a high computational cost. However, practical applications require a compact system.

Previously, Kobatake and coworkers [5–7] proposed a sharp-directional microphone-array system with a neural network. This system showed a sharper sensitivity than a linear system with the same size. Moreover, we achieved a sharper sensitivity than those reported in the previous studies [5,6] by optimizing the learning parameters for a neural network using a genetic algorithm (GA) [8].

Although the optimized system realized a sharp sensitivity for the learned frequency, the sensitivity was restricted to a narrow band around the learned frequency (i.e., 1,700 Hz in the previous studies [5–8]). This characteristic is unavoidable and is due to the structure of the previous system. Owing to the constructional limitation of the previous system, which will be described in detail in Sect. 2.2, the system cannot achieve uniformly sharp sensitivities at multiple frequencies. This becomes problematic if the microphone-array system is applied to signals with multiple frequency components. To resolve this issue, the previous system must be retrofitted.

In this study, we aim to achieve sharp sensitivities not only at a single frequency but also at multiple frequencies. We propose a new system in which the input structure of the previous system is modified. We also discuss the training method of the neural network because the method is important for obtaining a good performance. Furthermore, to evaluate the proposed system, the signal outputs of the system are analyzed for sinusoidal waves as well as amplitude or frequency modulated waves.

2. OVERVIEW OF THE PREVIOUS STUDIES

2.1. Structure of the System

Kobatake and coworkers [5–7] established a nonlinear-filter system consisting of a combination of $M+1$ microphones and a three-layered network (Fig. 1). This system was designed so that the desired signal is not fed into the
network when it arrives perpendicular to the array, which is called the look direction ($= 0^\circ$). Thus, this system can prevent the desired signal from being distorted. In other words, difference signals $d_0$ to $d_{M-1}$ do not contain the desired signal because it arrives at every sensor in-phase and the reference microphone (RM) output is subtracted from all the sensor microphone outputs.

Fixing the arrival direction of the target source to $0^\circ$ is not necessary. In this paper, we employ the simplest condition for which the target source is located perpendicular to the linear array in accordance with the previous studies [6–8]. If a look direction other than $0^\circ$ is required, it is easy to introduce appropriate delays for the microphones, which is well known as a phased array. When the look direction is not $0^\circ$, a beam pattern $G(\theta)$ as a function of the incident angle $\theta$ is transformed in the same way as for a linear array: when the look direction is $\theta_0$, the beam pattern is transformed to $G(\sin^{-1}(\sin(\theta) - \sin(\theta_0)))$. If $\theta_0$ is $0^\circ$, $G(\sin^{-1}(\sin(\theta) - \sin(\theta_0)))$ is identical to $G(\theta)$.

Neurons in the network use the following sigmoid function:

$$f(z) = \frac{2}{1 + \exp(-Tz)} - 1,$$

where $T$ is the temperature coefficient and $z$ is the instantaneous amplitude of the input signal.

### 2.2 Issue of the Previous System

Suppose that two signals with frequencies $f$ and $f'$ arrive from incident angles $\theta$ and $\theta'$, respectively. Kobatake and coworkers [5,6] indicated that the behavior of the system is identical when the following equation holds:

$$f \sin \theta = f' \sin \theta'.$$

Thus, if the sensitivity of the system is optimized for frequency $f$, it is degraded for another frequency $f'$.

Figure 2 shows the output level sensitivity of the previous system [8], which was optimized to a sinusoidal wave of 1,700 Hz. The sensitivity was measured under the condition that an input signal arrives from each direction independently (from $0^\circ$ to $90^\circ$). The bold solid, thin solid, and dashed lines represent the sensitivities for an input sinusoidal wave of 1,700 Hz, 850 Hz, and 425 Hz, respectively. The sensitivities at 425 Hz and 850 Hz are calculated using Eq. (2). The sensitivities at 425 Hz and 850 Hz are inferior to that at 1,700 Hz. This issue is due to the input pattern to the network, which is a pattern of the spatial distribution of sound pressures that are observed with the microphones. Owing to this constructional limitation, an input pattern having a different frequency from the trained frequency is recognized as an input pattern with the trained frequency but arriving from a different incident angle on the basis of Eq. (2). In other words, the previous system [6,8] cannot be trained with multiple frequencies.

This issue must be resolved to achieve a wide-band sharp sensitivity. However, simply optimizing the learning parameters of the network [8] does not resolve this issue.

### 3. Proposed System

On the basis of the above-mentioned issue and the notion of a time-delay neural network [9], we propose an expanded system where the spatial pattern, which is delayed by as much as one sampling period, is also treated as part of the input pattern (Fig. 3). In other words, the network learns the temporal-spatial patterns that are combinations of numerical values of network inputs at certain discrete times $n$ and $n - 1$. The input signals of the proposed system provide the information that the frequencies of two signals are different even when Eq. (2) holds; although the two signals show an identical spatial pattern at a certain discrete time, they exhibit different spatial patterns at the next discrete time.

This system assumes a straight-line array in accordance with the previous studies [6,8]. When the sound source is far from the microphone array, the arriving signals can be
where it is assumed that a desired signal arrives perpendicular to the array, which is called the look direction (\(=0^\circ\)). (This condition is the same as that in the previous studies [6,8].) The interference noises \(u_n(t)\) and \(u_{RM}(t)\), which are observed at the \(m\)-th microphone and RM, respectively, are expressed by Eqs. (3) and (4). Here the equations are denoted in an analog manner because the time delays are analog values.

\[
u_m(t) = u(t - \tau_m) \quad (3)
\]
\[
u_{RM}(t) = u(t - \tau_{RM}) \quad (4)
\]

\(\tau_m\) and \(\tau_{RM}\) denote the delays that occur when the noise originates from a direction other than the look direction.

The sampled output \(x_m(n)\) of the \(m\)-th microphone, which consists of the desired signal \(s(n)\) and \(u_m(n)\), is expressed in a digital manner with a discrete time \(n\) in Eqs. (5) and (6).

\[
x_m(n) = s(n) + u_m(n), \quad m = 0, 1, \cdots, M - 1 \quad (5)
\]
\[
x_{RM}(n) = s(n) + u_{RM}(n) \quad (6)
\]

by combining Eqs. (5) and (6), the input to the network, \(d_m(n)\), is given by

\[
d_m(n) = x_m(n) - x_{RM}(n), \quad m = 0, 1, \cdots, M - 1 \quad (7)
\]

where \(d_m(n)\) is the difference between the signal of each microphone and that of the RM.

This system uses these difference signals as the input to the neural network. The neural network is organized by training so that its output gives the noise signal received by the RM, \(u_{RM}(n)\). The noise signal estimated by the network is subtracted from the RM output \(x_{RM}(n)\) to restore the desired signal \(s(n)\).

The proposed system uses delayed signals. The expansion where \(d_m(n-1)\) is also used as part of the network input allows two signals with different frequencies to be learned simultaneously regardless of whether Eq. (2) holds. Additionally, to achieve a wide-band sensitivity, we plan to use multiple frequencies in the network training process.

4. EVALUATION EXPERIMENTS

4.1. Experimental Setup

4.1.1. Conditions

The basic conditions of the microphone arrangement and the training method of the system are equivalent to those described in the previous studies [5,8]. Namely, seven microphones are arranged on a straight line in 0.1 m intervals, including the RM. Network training is executed by back-propagation.

The sound sources are simulated under the following conditions:

- The acoustic wave is regarded as a plane wave because the sound source is distant.
- The sound velocity is 340 m/s.
- The experiment is simulated without acoustic reflections, assuming a free-field condition.
- The training input signal is a sine wave with a frequency of 425 Hz, 850 Hz, or 1,700 Hz, where the amplitude of the wave is normalized to 1. In accordance with the previous studies [7,8], the frequency of 1,700 Hz is selected for the simulation. The frequencies of 425 Hz and 850 Hz are one and two octaves lower than 1,700 Hz, respectively. (Although selecting other frequencies and increasing the number of trained frequencies would be possible, these three frequencies are selected for our trial.)
- The sampling frequency is 8 kHz.

The input layer of the network contains 12 neurons because, in addition to the signals of the six microphones without a delay, the delayed signals are treated as inputs. The optimization, which is described in the next section, is used to determine the number of neurons in the hidden (middle) layer.

The training process is executed under the following conditions:

- A sinusoidal wave arrives as an input wave to the system. The training data for this system include the above three frequencies because the frequency of the input wave is randomly selected from the three frequencies. As a result, a single set of weights is obtained through the training.
- The amplitude of a wave is 1. Consequently, the maximum amplitude of difference signals is 2, which occurs when the outputs from the RM and a microphone are out of phase due to a time delay.
- The arrival direction of the input wave shifts from 0° (the look direction) to 90°. The interval of the arrival directions between 0° and 9° is set to 1° (0, 1, ∙∙∙, 9°) to achieve uniformly sharp sensitivities at multiple
frequencies. When the arrival direction exceeds 10°, the interval is set to 5° (10, 15, · · · , 90°).

- Training patterns are prepared for every arrival direction. Each training pattern consists of 14 instantaneous values for the sound pressure of the arriving signal; seven values are the sound pressures observed at the seven microphones at a discrete time \(n_0\), and the other seven are those at \(n_0 - 1\). To train the network, a teacher signal is determined for each training pattern.

- The teacher signal for the network output is zero when a sound arrives from the look direction (0°) because all of the difference signals and their delayed signals input to the network are zero in this case.

- When the input sinusoidal wave arrives from an angle of 6 to 90°, it is regarded as a noise because we aim to suppress the output of the system. In this case, the teacher signal for the network output is the instantaneous value of the noise in the RM output at a discrete time \(n_0(\text{RM}(n_0)); \) on the basis of the input pattern, the network recalls the noise signal that must be subtracted from the RM output. Because the network is trained to recall the noise signal in the RM output, the output from the network is the noise signal to be subtracted.

- When the input wave arrives from an angle of 1 to 5°, it is regarded as the desired signal \((\sigma(n))\). However, the output from the system is attenuated as the incident angle is shifted from 1 to 5° because the output must be suppressed at 6°. In other words, as the arrival direction shifts from 0 to 6°, the neural network output shifts from zero to the instantaneous value of the RM output \((\text{RM}(n_0)\text{ at } 6°)\). In order to smooth the shift of the network output between 1° and 5°, the teacher signals for the network output are set to the following values: 0.042 \(\text{RM}(n_0)\) for 1°, 0.150 \(\text{RM}(n_0)\) for 2°, 0.320 \(\text{RM}(n_0)\) for 3°, 0.523 \(\text{RM}(n_0)\) for 4°, and 0.713 \(\text{RM}(n_0)\) for 5°. These target values are determined by referring to the sensitivity obtained in our previous study [8], in which the network was trained using only a 1,700 Hz sinusoidal wave.

- On the basis of the fact that one period of the 425 Hz signal corresponds to about 20 sampling intervals when the sampling frequency is 8 kHz, 20 training patterns are prepared for each frequency and each incident angle by shifting \(n_0\). Similarly, 20 patterns are used for the 850 Hz signal. However, the number of patterns is increased to 40 for the 1,700 Hz signal (the reason for this will be described in Sect. 5.1.2).

### 4.1.2 Parameter optimization of the network

Prior to executing the training process of the network, the values of six parameters must be set: the number of neurons in the hidden layer \((N_{\text{hid}})\), a training parameter for back-propagation \((\eta)\), the maximum values of the absolute values to randomly set default values of the weights between the input-hidden layers and the hidden-output layers \((WR_{ih} \text{ and } WR_{ho}, \text{ respectively})\), the temperature coefficient \((T)\) used in Eq. (1), and the minimum value of mean squared error (MSE) \((\text{MSE}_{\text{min}})\) at frequency \(f\) when the input signal arrives from the \(d^o\) direction. Thus, the cost function used in the training of the network is as follows:

\[
\text{MSE} = \frac{1}{F \cdot D \cdot N} \sum_{f=1}^{F} \sum_{d=0}^{D} \sum_{n=0}^{N-1} (s_{gf}(n) - o_{gf}(n))^2, \tag{8}
\]

where \(F\) is the number of the training frequency indices \(f\). \(D\) is the number of the arrival directions \(d^o\), which are increased in 1° steps. The values of \(F\) and \(D\) are 3 and 90, respectively, as described in Sect. 4.1.1. \(N\) is the number of the temporal-spatial patterns for each frequency. The value of \(N\) is 20 or 40 as described in Sect. 4.1.1.

Herein, a GA [10] is employed to optimize the parameter values for the proposed system in the same manner as in the previous study [8].

Table 1 shows the conditions of the GA, where selection, crossover, and mutation are genetic operations. The crossover rate and mutation rate are parameters of these genetic operations. Each chromosome consists of a combination of these six parameters, which are randomly set within the search range shown in Table 2 at the beginning of the training process.

The evaluation function, which is a parameter used to evaluate the optimality (score) in the GA, is determined as Eq. (9), similarly in our previous study [8].

\[
\text{Evaluation function } E = \frac{O_d}{\sum_{d=5}^{90} O_d}, \tag{9}
\]

\(O_d\) is the sum of the powers of the output signal of input signals arriving from \(d^o\) directions. The sum of the values of Eq. (9) for the three trained frequencies is used as the final evaluation function.

Table 2 shows the optimized values of the parameters for the network and its training conditions.

### 4.1.3 Weight value optimization of the network

After optimizing the network parameters, 500 sets of default weight values are randomly set using the optimized

---

**Table 1** Conditions of GA.

| Selection            | Roulette wheel selection |
|----------------------|--------------------------|
| Crossover            | Single point crossover   |
| Crossover rate       | 90%                      |
| Mutation rate        | 2%                       |
| Number of chromosomes| 100                      |
| in a generation       | 100                      |

---

**Table 2** Optimized values of the parameters.

| Parameter          | Value          |
|--------------------|----------------|
| Number of neurons  |                |
| Training parameter |                |
| Back-propagation   |                |
| Temperature        |                |
| Minimum MSE        |                |

The values of $WR_{ih}$ and $WR_{ho}$. The training process with each set is executed using the optimized parameters for the network, and the set of weights with the sharpest sensitivity is selected as the optimized set of weights, that is, the sensitivity with the highest signal/noise ratio between the look direction (0°/C14) and the other directions (5–90°/C14) for 425–1,700 Hz is selected. The network consisting of this set of weights after the training is used in the following evaluation.

### 4.2. Results

#### 4.2.1. Evaluation using sinusoidal signals

After completing the training process using the optimized parameters, an experiment (computational simulation) on the proposed system was executed using sinusoidal waves with the frequencies used in the training. Figure 4 shows the sensitivity of the proposed system (Fig. 3) with the optimized network when an input signal arrives from each direction independently (from 0 to 90°). The sound source conditions are the same as those described in Sect. 4.1.1. The sharp sensitivities with a five-degree main lobe and side lobes suppressed by about 20 dB or more for the three frequencies indicate that the optimization is successful. The suppression seems to be larger at higher frequencies.

To observe the sensitivity of sinusoidal signals with frequencies other than the three trained frequencies, the system is also tested using sinusoidal signals with frequencies of 100 to 1,700 Hz in 10 Hz steps independently (Fig. 5). Sharp sensitivities are achieved for frequencies above 425 Hz. Similar to the trained signals, suppression is greater at higher frequencies. The sensitivities with a half width of 5° are produced at each frequency from 425 to 1,700 Hz, demonstrating that the proposed system successfully yields a sharp sensitivity for sinusoidal signals with the multiple frequencies.

The results suggest that the system with the neural network is valid for trained input patterns as well as patterns similar to the trained patterns. In other words, the network can interpolate the output signal of inputs that include untrained patterns at untrained frequencies among the trained frequencies.
4.2.2. Examination using AM or FM wave

The characteristics of the system were examined by employing narrow band complex signals such as an amplitude-modulated (AM), or frequency-modulated (FM) wave as the input signal. To synthesize AM (FM) signals, the carrier was an 850 Hz sinusoidal wave modulated in amplitude (frequency) by a 30 Hz sine wave [8]. The solid and dotted lines in Fig. 6 represent the sensitivities for input signals of a 15% AM wave and a 15% FM wave arriving, respectively. Each result maintains a sensitivity that is as sharp as the result for 850 Hz in Fig. 4. In our previous study [8], when an AM or FM wave was the input, the sensitivity deteriorated compared with that of a sinusoidal wave input. Therefore, the proposed system achieves improved results compared with the previous system [8].

5. DISCUSSION

5.1. Evaluation Using Sinusoidal Signals

Figure 5 shows that uniformly sharp sensitivities are obtained at multiple frequencies between 425 and 1,700 Hz. This achievement is attributed to the following three factors:

1. Optimization of the network parameters and its training parameters.
2. Preparation of sufficient training data.
3. Selection of default values of weights to avoid partial optimization.

5.1.1. Optimization

Optimization of the parameters of the network structure ($N_{hid}$ and $T$) and the training ($\eta$, $WR_{lb}$, $WR_{hb}$, and $MSE_{min}$) helps maximize the potential of the system in order to achieve a sharp sensitivity. In this study, the input structure of the previous system is modified to allow the proposed system to handle multiple frequencies, while the optimization is employed to yield a sharp beam pattern.

The beam pattern for the trained frequency of 1,700 Hz (shown in Fig. 4) is sharper than the one reported in the previous study [8]. The optimization of such parameters effectively achieves sharp sensitivities at trained frequencies, including suppression of the side lobes and sharpening of the main lobe. However, the sensitivity at 1,700 Hz obtained by the proposed system, is slightly less than that reported in our previous study [8], suggesting that increasing the number of training frequencies may degrade the sensitivity compared with training for a single frequency. This is attributed to the fact that the network trained using sinusoidal signals at the multiple frequencies is not tuned to a specific frequency.

5.1.2. Examination of training data

Even if fixed parameter values in the training network are used, the results of trained networks depend on the data used in the training process. Figure 7 [11] shows the results obtained via a preliminary training process using different training data from those described in Sect. 4.1.1.

The sensitivities at the untrained frequencies (e.g. around 1,700 Hz) in Fig. 7 are inferior to those in Fig. 5. In the case of training to obtain the result in Fig. 7, the training data for each frequency and each incident angle consisting of 20 temporal-spatial patterns, which do not seem to be sufficient for the network to learn sinusoidal signals at high frequencies. The temporal-spatial patterns at high frequencies are more complex than those at low frequencies; that is, it is difficult for the network to learn the regularity of a sinusoidal wave at high frequency. Therefore, untrained temporal-spatial patterns at high frequencies of around 1,700 Hz are prone to be uncontrollable for the network.

On the basis of the above assumption, to achieve the results shown in Fig. 5, the training data are modified; the number of training patterns for 1,700 Hz is increased from...
20 to 40 patterns as described in Sect. 4.1.1. This allows the network to learn the temporal-spatial pattern for a frequency of 1,700 Hz more precisely, which enables the network to control the sinusoidal waves at high frequencies of around 1,700 Hz.

5.1.3. Effects of default values of weights

Even if the same training parameter values and the same training data are used, the results still depend on the default values of the weights, which are randomly set at the start of the training process on the basis of the parameters $WR_b$ and $WR_{ih}$.

Figure 5 has the best sensitivities selected from the results of multiple trainings as described in Sect. 4.1.3. The training processes are executed using each of the 500 sets of the default values of weights.

Depending on the default values of weights, the network may lose versatility for wide-band frequencies due to overtraining or partial optimization for the three trained frequencies (425, 850, and 1,700 Hz). However, as shown in Fig. 5, uniformly sharp sensitivities between the trained frequencies are achieved and a highly versatile system is obtained by selecting the best result among numerous default sets as described above.

In the future, an evaluation of the sensitivities for untrained frequencies should be added to the evaluation function used to optimize the GA parameters to prevent the training results from overtraining or partial optimization to the training data used. Moreover, increasing the number of trained frequencies may enhance the versatility of the system.

5.2. Examination Using AM or FM Wave

In terms of the sensitivity for the AM wave shown in Fig. 6, the beam pattern is superior to that reported in the previous study [8], which employed a single frequency in the training process. In other words, the proposed system can be used for complex signals such as AM waves. This improvement is likely due to the increased number of signal patterns used to train the network at multiple frequencies compared with in our previous study [8].

As the depth or degree of FM or AM modulation increases or as the modulation frequency increases, the performance of the proposed system deteriorates. This behavior was also observed in the previous system [6,8]. As considered previously [8], this behavior attributed to the sensitivities of the system, which depend on the similarity of the waveforms between the input wave and the training wave. In this study, the training waves are sinusoidal. In the case of an AM or FM wave with a high modulation degree or frequency, its waveform differs from a sinusoidal wave, leading to discrepancies in the temporal-spatial patterns. These untrained patterns are considered to be responsible for the observed deterioration.

The system may be hardly applicable to wide-band complex signals such as speech and white noise. If an already known sound (waveform) is used in the training process, the proposed system should maintain sensitivities to the sound. However, this idea is impractical since a microphone is generally used to record unknown sounds. In the future, the system should be improved so that it can maintain its sharp sensitivity for any input wave. For example, training with various waveforms would be effective to enhance the versatility of the system.

6. CONCLUSION

After reviewing the previously reported system [5–8], we propose a system with a new input structure. The proposed system achieves a sharp sensitivity not only at the three trained frequencies but also at untrained frequencies in the range of the training frequencies (i.e., 425–1,700 Hz). Moreover, the proposed system maintains a sharper sensitivity for AM/FM wave inputs than the previously reported system [8].

In the future, the neural-network-based microphone-array system should be improved in terms of the following aspects. (1) To handle wide-band signals, the intervals between microphones should be reduced because an interval of less than 0.042 m can handle a frequency of up to 4,000 Hz, which is applicable for telephone-speech sound. Trained frequencies as well as the number of frequencies should be selected depending on the intervals. (2) The proposed system can currently handle two sounds simultaneously (one signal arriving from the look direction and one noise arriving from another direction). However for practical applications, the system must be expanded to handle multiple noise sources simultaneously. (3) The proposed system was examined assuming a free-field condition in this study. The system should be examined under a reverberation condition. (4) Applying deep-learning to the network should improve the stability and versatility of the system.

ACKNOWLEDGMENTS

This study was partially supported by the Japan Society for the Promotion of Science (KAKENHI, Nos. 24560519, JP16K06384).

REFERENCES

[1] J. L. Flanagan, J. D. Johnston, R. Zahn and G. W. Elko, “Computer-steered microphone arrays for sound transduction in large rooms,” J. Acoust. Soc. Am., 78, 1508–1518 (1985).
[2] M. Omologo, M. Matassoni, P. Svaizer and D. Giuliani, “Microphone array based speech recognition with different talker-array positions,” Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 97), Vol. 1, pp. 227–230 (1997).
[3] Y. Kaneda and J. Ogaha, “Adaptive microphone-array system for noise reduction,” IEEE Trans. Acoust. Speech Signal Process., ASSP-34, 1391–1400 (1986).
[4] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” IEEE Trans. Antennas Propag., AP-23, 27–34 (1982).

[5] H. Kobatake, W. Morita and Y. Yano, “Super directive sensor array with neural network structure,” Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP 92), Vol. 2, pp. 321–324 (1992).

[6] W. Morita and H. Kobatake, “Super directive microphone array system with nonlinear multi-layer structure,” J. Acoust. Soc. Jpn. (J), 49, 28–33 (1993) (in Japanese).

[7] W. Morita and H. Kobatake, “Analysis of nonlinear super directive microphone array system,” J. Acoust. Soc. Jpn. (J), 49, 34–39 (1993) (in Japanese).

[8] A. Iseki, Y. Kinoshita and K. Ozawa, “Optimization of neural-network-based superdirective microphone-array system using a genetic algorithm,” Acoust. Sci. & Tech., 36, 326–332 (2015).

[9] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano and K. J. Lang, “Phoneme recognition using time-delay neural networks,” IEEE Trans. Acoust. Speech Signal Process., 37, 328–339 (1989).

[10] J. H. Holland, Adaptation in Natural and Artificial Systems (The University of Michigan Press, Ann Arbor, 1975).

[11] A. Iseki, K. Ozawa and Y. Kinoshita, “Neural network-based microphone array learning of temporal-spatial patterns of input signals,” Proc. IEEE 3rd Global Conf. Consumer Electronics (GCCE 2014), pp. 88–89 (2014).