Optimization of neural-network-based superdirective microphone-array system using a genetic algorithm

Akihiro Iseki*, Yuichiro Kinoshita† and Kenji Ozawa‡

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

(Received 27 June 2014, Accepted for publication 6 January 2015)

Abstract: A previous study indicated that superdirectivity is achieved with a microphone-array system consisting of seven microphones and a neural network. In this work, we aim to improve the directivity by optimizing the parameters of the neural network using a genetic algorithm. A computational simulation shows that the optimized system can produce superdirectivity with a half-width of five degrees. The optimized system improves the directivity because its side-lobes are suppressed by more than 10 dB compared with the results of the previous study. This suppression is also observed in terms of harmonic distortions. Moreover, in an examination using AM and FM waves as input signals, the optimized system achieves higher performances than those in the previous study.

Keywords: Microphone array, Neural network, Genetic algorithm, Superdirectivity

PACS number: 43.60.Fg, 43.60.Np [doi:10.1250/ast.36.326]

1. INTRODUCTION

We aim to produce a microphone-array system exhibiting superdirectivity because such a system can effectively reduce noise in diverse situations such as recording a target sound in a noisy meeting and allowing a robot to localize sound sources [1]. However, the microphone array must be small and compact in order to be applicable in a variety of practical situations.

To date, delay-and-sum arrays [2,3] and adaptive microphone arrays have been proposed [4,5] to achieve sharp directivity. However, their performances are limited because they are linear systems that require a number of microphones and high computational costs. To attain a high performance with a compact system, we focus on a nonlinear structure because a nonlinear system has the potential to achieve a sharper directivity than a linear system.

In previous studies, Morita and coworkers [6–8] proposed a sharp directional microphone-array system with a multilayer structure, namely, a neural network. Although their system achieved sharp directivity for certain frequencies, it is unclear if their system realized the best performance of its potential directivity; that is, it is uncertain whether the network parameters were optimized. We assume that an array system can achieve a sharper directivity than reported in previous studies [6,8].

In this study, we aim to improve the directivity using a genetic algorithm (GA) to optimize the parameters of the network in the microphone-array system. Furthermore, to verify the improved directivity, the signal outputs and harmonic distortions of the optimized system as well as the outputs for the AM and FM waves are analyzed.

2. OVERVIEW OF THE PREVIOUS STUDY

2.1. Linear-Filter System

The simplest example of a microphone-array system is a summing array system, which is a linear system. Its directivity is produced by summing the outputs of linearly arranged microphones through a linear filter. The simple model of the signal received by a microphone array and its output are shown as:

$$y(t) = \sum_{m=0}^{M-1} w_m x_m(t),$$  \hspace{1cm} (1)

where $x_m(t)$ is the input signal received at the $m$-th microphone at time $t$, and $M$ is the number of microphones. The weight $w_m$ is associated with each $m$-th microphone.

2.2. Nonlinear-Filter System

Morita and coworkers [6–8] established a nonlinear-filter system consisting of a combination of six micro-

*e-mail: akihirochristeve@hotmail.com
†e-mail: ykinoshita@yamanashi.ac.jp
‡e-mail: ozawa@yamanashi.ac.jp
phones arrayed at 0.1 m intervals and a three-layered network (Fig. 1). Because this network is nonlinear, the principle of superposition does not hold for this system. The idea of this system is as follows. In previous works [6–8], severe nonlinear distortions were observed in the desired signal when the sum of the desired signal and noise were fed into the neural network owing to the nonlinear characteristics of the sigmoid function. The system in Fig. 1 is designed so that the desired signals, which arrive from a direction perpendicular to the array, called the look direction (= 0°), are not fed into the network. Furthermore, this system can prevent the desired signals from being distorted even if the microphones receive the sum of the desired signal and noise. The difference of the outputs between each microphone and the reference microphone (RM) does not contain the desired signal components because the desired signals arrive at each sensor in-phase. This system handles sound signals that vary between plus and minus around zero. Neurons in the network, therefore, use the following sigmoid function:

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

where \( T \) is a temperature coefficient. This nonlinear system is aimed at avoiding nonlinear distortions, which deteriorate the system characteristics. This system is designed on the basis of the idea that a target signal coming from the look direction is output without nonlinear processing, while noise is suppressed with nonlinear processing. In this system, a straight-line array is assumed. The \( M \) microphones and an extra sensor, which is called RM, are arrayed on the same straight line. The sound source is far from the microphone array so that the arriving signals can be regarded as a plane wave. The output \( x_m(t) \) of the \( m \)-th microphone, which consists of the desired signal \( s(t) \) and interference noise \( u(t) \), is expressed as

\[
x_m(t) = s(t) + u(t - \tau_m), \quad m = 0, 1, \cdots, M - 1 \tag{3}
\]

\[
x_{RM}(t) = s(t) + u(t - \tau_{RM}), \tag{4}
\]

where \( \tau_m \) denotes a delay that occurs when sound comes from a direction other than the look direction. By combining Eqs. (3) and (4), the input to the network, \( d_m(t) \), is

\[
d_m(t) = x_m(t) - x_{RM}(t) = u(t - \tau_m) - u(t - \tau_{RM}), \quad m = 0, 1, \cdots, M - 1 \tag{5}
\]

where \( M \) is the number of microphones other than RM, and \( d_m(t) \) is the difference between the signal of each microphone and that of RM. As an important point, \( d_m(t) \) becomes zero when the sound from the look direction arrives in phase at each microphone in the array; that is, no signals are input into the network. Instead of using the sensor output itself, this system uses these difference signals as the input to the neural network. The neural network is organized by training such that its output gives the noise signal received by RM, \( u(t - \tau_{RM}) \). The noise signal is subtracted from the RM output \( x_{RM}(t) \) to restore the desired signal \( s(t) \).

3. PARAMETER OPTIMIZATION OF THE NETWORK

3.1. Training Method of Network and Experimental Conditions

To compare to the results of this study, the nonlinear system developed by Morita et al. [6,8] is initially simulated. In this section, we describe the conditions of the computational simulation. The conditions of the structure and coefficients are equivalent to those described in the previous study [8]. Namely, six microphones are arranged on a straight line in 0.1 m intervals, and the RM is placed on the same line next to the sixth microphone with a 0.1 m interval. Training of the network is executed under the condition of a single sound by the backpropagation method. If a sound arrives from the look direction, the teacher signal for the network output is zero, but the teacher signal is the RM output when the sound arrives from any other direction. Thus, the nonlinear system subtracts noise from the RM signal, whereas the network forms this noise from the differences among signals (\( d_m \) in Fig. 1).

Sound sources are simulated under the following conditions:

- The acoustic wave is handled as a plane wave because the sound source is distant.
- The sound velocity is 340 m/s.
- The input signal for training is a sine wave with a frequency of 1,700 Hz. The interval of each microphone is half its wavelength (0.1 m).
- Two sound waves arrive simultaneously (the target signal from the look direction and noise from another direction). Both the signal and noise are sine waves with normalized amplitudes that vibrate from −1 to +1.
Table 1 Ranges of parameters of the neural network.

| Parameter | Range       |
|-----------|-------------|
| $\eta$    | 0.001 to 0.1|
| $WR_{ih}$ | 0.1 to 100  |
| $WR_{ho}$ | 0.1 to 100  |
| $WR_b$    | 0.1 to 5    |
| $T$       | 0.01 to 10  |
| $MSE_{\text{min}}$ | 0.001 to 0.1 |

There are six neurons in the input layer of the network because this is equivalent to the number of microphones. Additionally, there are six neurons in the hidden (middle) layer. The training processes are executed as the arrival direction of noise shifts in five-degree intervals (0°, 5°, ⋯, 90°).

3.2. Optimization Method

In order to execute the training process of the network, the values of several parameters must be set for the network. However, these values were not described in the previous study [8]. Although optimizing the parameter values in the network is important for achieving a sharper directivity than that in the previous study, the optimal values are unclear. Herein, a GA [9] is employed to optimize the values of the parameters.

GA is employed for multiple reasons. First, optimization of the parameter values in the network incurs a large calculation cost, which can be effectively reduced by GA. That is, the training process of the network requires a large calculation cost because it needs to continue until Mean Squared Error (MSE) is reached to a certain target value. Therefore, to search for the optimum values of many parameters, a number of training processes with each parameter value must be executed. GA decreases the entire calculation cost by reducing the number of searches for the optimum parameters of the network. Second, estimating the proper values for the parameters is difficult owing to the lack of preliminary information. Consequently, the search range is wide. GA can accommodate such a wide search range within a short searching time. Third, although the optimum parameters need not necessarily be obtained since we aim to improve the performance of a nonlinear system, GA can determine parameter values, that are close to the optimum values.

Table 1 shows the parameters for the network, where $\eta$ represents a training parameter for backpropagation. $WR_{ih}$, $WR_{ho}$, and $WR_b$ are the maximum values of the absolute values to randomly set default values of weights for the input layer to the middle layer, the middle layer to the output layer, and the bias parameter, respectively. (The default values of these weights are set randomly between plus and minus where the range is shown as the absolute values in Table 1.) $T$ is the temperature coefficient used in Eq. (2), while $MSE_{\text{min}}$ is the minimum value of MSE between the desired signal $s(t)$ and the output of the system (if MSE is less than or equal to $MSE_{\text{min}}$, the training iteration stops).

Table 2 shows the conditions of the GA, where selection, crossover, and mutation are genetic operations. The crossover rate and the mutation rate are the parameters for the genetic operations. Each chromosome consists of a combination of the six parameters which are randomly set within the range shown in Table 1. The GA optimizes the parameters for the network by obtaining high-scoring chromosomes for each generation, which is an iterative process. Therefore, a total of 10,000 (100 × 100) of searches of high-scoring chromosomes are conducted to optimize the parameters for the network.

The evaluation function, which is a parameter to evaluate the optimality (score) in GA, is determined as Eq. (6); this equation strives to “maintain a sharp directivity for the look direction (= 0°), while suppressing the output of input from other directions such as 5° to 90°.” In the previous study [6], the training processes were executed as the arrival direction of noise was shifted in five-degree intervals (0°, 5°, ⋯, 90°) while the look direction was only 0°. On the basis of the above process, Eq. (6) is employed with the look direction of 0°, while the other directions are more than 5°.

Evaluation function $E = \frac{O_h}{\sum_{d=5}^{90} O_d}$ (6)

Where, $O_d$ is the sum of the powers of the output signals of input signals arriving from $d°$ directions.

3.3. Optimized Parameters

Table 3 shows the parameters optimized by the GA. Additionally, the GA allows quasi-optimized parameters to be realized. For comparison, Table 4 shows the quasi-optimized parameters.
4. IMPROVEMENT OF DIRECTIVITY
OF NONLINEAR SYSTEM

4.1. Beam Patterns of System with Optimized Parameters

Figure 2 shows the output waveform of the system shown in Fig. 1 with the optimized parameters shown in Table 3 for the network, when a 1,700 Hz sinusoidal signal with an amplitude of 1 arrives from 0° (the look direction) or 5° (another direction). The sampling rate is 44.1 kHz. This simulation is conducted under the same condition as described in Sect. 3.1. The system outputs signals of input arriving from the look direction without nonlinear processing. Additionally, the noise arriving from the other directions (e.g., 6° to 90°) is suppressed (Fig. 2(b)).

Figure 3 shows the beam patterns of this system when an input signal arrives from each direction independently (from 0° to 90°), i.e., the number of sound sources is one. The solid, dashed, and dotted lines represent the characteristics of the system using the optimized parameters shown in Table 3, those from a previous study [8], and those of a linear system, respectively.

The characteristics of the optimized network indicate almost the same main lobe as that of the previous system, which are sharper than those of the linear system. Furthermore, it is observed that the beam patterns of this system are greatly improved relative to those in the previous study; the pattern in this system maintains a sharp main lobe while the output levels of the side-lobes are 15–20 dB lower than those in the previous study.

4.2. Examination of Harmonic Distortions

To examine the system performance with the optimized parameters of the network, harmonic distortions are analyzed by comparison with those of the previous study. Figure 4 shows the harmonic distortions of the system output under the following conditions: (a) a 1,700 Hz sinusoidal signal arriving from 0°, the look direction, and (b) a signal arriving from 10°. The solid and dotted lines represent the results of this study and the previous study [8], respectively. The optimized system does not output a
harmonic distortion for the signal from the look direction (Fig. 4(a)). Additionally, the output of the optimized system suppresses harmonic distortions for the non-look directions (Fig. 4(b)). Thus, compared with the previous study, the system with optimized parameters inhibits harmonic distortions.

4.3. Examination Using AM or FM Waves

The characteristics are examined by employing an AM (amplitude-modulated) or FM (frequency-modulated) wave as the input signal. To synthesize AM (FM) signals, a 1,700 Hz sine wave is used as a carrier and is modulated in amplitude (frequency) by a 30 Hz sine wave. The solid lines in Fig. 5 represent the beam patterns for an input signal arriving from each direction independently under the following conditions: (a) the beam patterns for an input of a 15% AM wave, employing the optimized parameters for the network, and (b) the beam patterns for an input of a 15% FM wave, employing the quasi-optimized parameters. For comparison, the dotted line shows the beam patterns obtained in the previous study [8].

Although the beam patterns for directions other than the look direction are slightly deteriorated, the main lobe in Fig. 5 is as sharp as the one in Fig. 3. For both AM and FM waves, parameter optimization improves the beam patterns compared to the previous study [8].

5. DISCUSSION

5.1. Parameter Optimization Using GA

The parameters selected by the GA are 0.013 for $\eta$, and sub-equal values (0.161 and 0.253) for $T$, while $WR_{ih}$, $WR_{ho}$, and $WR_b$ differ between the optimized parameters and quasi-optimized parameters (Tables 3 and 4). These weight parameters determine the range of the initial values of the weight coefficients, which are set randomly in the training process of the network. Therefore, the weight coefficients are less affected by the parameters $WR_{ih}$, $WR_{ho}$, and $WR_b$, and such parameters may be too ambiguous to use as optimization parameters. Additionally, the actual bias parameters converge to almost zero in the training process of the network. This phenomenon has been noted in a previous study [7,8]. Thus, which parameters are selected to be optimized should be further investigated.
5.2. Examination of Harmonic Distortions

In terms of harmonic distortions, a system with optimized parameters exhibits an improved performance compared with that in the previous study [6,8] (Fig. 4). This improvement is attributed to the bias parameter. The optimized parameters do not include the bias parameter because the optimization results indicates that the bias parameter converges to zero. In the previous study [8], the authors assumed that their system possessed the bias parameter.

Additionally, the output levels of the optimized system are mostly improved compared with those in the previous study. However, 5,100 Hz (third-order harmonic) and 8,500 Hz (fifth-order harmonic) distortions of the optimized system are almost the same as those in the previous study. Thus, the optimization also effectively suppresses harmonic distortions.

5.3. Examination Using AM and FM Waves

Quasi-optimized parameters are employed to examine the beam patterns for the FM waves, because the beam patterns deteriorate slightly when optimized parameters are employed, compared with the results of using the quasi-optimized parameters (Fig. 5(b)). This deterioration is attributed to the fact that FM waves are not used in the training process.

When using AM and FM waves with a 15% modulation depth and 30 Hz modulation frequency, the main lobe (0–5°) maintains its sharpness. However, as the depth or degree of modulation of the FM or AM wave increases or the modulating frequency becomes higher, the beam patterns for the AM and FM waves deteriorate probably because the difference in the beam patterns between low and high degrees of modulation of the AM or FM wave depends on the similarity of waveforms between the input wave and trained wave (which is a 1,700 Hz sine wave in this paper). In the case of an AM or FM wave with a high modulation degree or frequency, its waveform differs from a 1,700 Hz sine wave, which leads to different combinations of numerical values of network inputs at a certain time t. These untrained combinations likely cause the deteriorations.

However, the sharp main lobe is maintained even if the waveforms of the input signals differ from the sine wave used in the training process. This is most likely because the inputs of the AM or FM waves to the network are similar to those used in the training process (sinusoidal signal) when the arrival directions are close to the look direction. As a result of the structure of this system (Fig. 1), the smaller the absolute values of the input to the network, the closer the arrival direction is to the look direction. Therefore, it is assumed that the network has been trained to output small values when the absolute values of the inputs are small. This tendency is maintained even if the waveforms of the input AM or FM wave signals slightly differ from those used in the training process.

Additionally, the beam patterns for the AM or FM waves are improved compared with those of the previous study. Hence, optimization of the network parameters is still effective for maintaining the directivity even if the waveform of the input signal differs from those of the trained signals.

In this study, sinusoidal, AM, and FM waves were used as input signals to evaluate the effect of the optimization by comparison of our result with those of the previous study [6] in which the same signals were used. According to the result of this work, although the potential of the array system using a neural network was brought out by parameter optimization, it is obvious that this system cannot perform correctly with untrained input signals. In the future, the system should be improved so that it can maintain its sharp directivity for any input wave, even if the waveform differs from those used in the training process. Moreover, the structure and training process should be improved so that the array system may retain the sharp directivity for broad-band sounds such as speech and white noise.

6. CONCLUSION

After reviewing the nonlinear system developed in a previous study [6-8], the beam characteristics were improved by optimizing the network parameters via a GA. The optimized results not only showed a sharp directivity, namely superdirectivity, but the characteristics of the side-lobes were also improved by 15–20 dB when compared with those in a previous study [6,8]. Moreover, such improved characteristics were also observed when using AM or FM waves as input. Thus, optimization of neural network via a GA improved the performance of the microphone-array system in this study.

In the future, this method should be expanded to handle plural frequencies (wide bandwidths) in order to realize a sharp directivity for a voice band, for example.

ACKNOWLEDGMENTS

This study was partially supported by the Japan Society for the Promotion of Science (KAKENHI, No. 24560519).

REFERENCES

[1] K. Nakadai, T. Laurens and H. G. Okuno, “Active audition for humanoid,” Proc. 17th Natl. Conf. Artificial Intelligence (AAAI-2000), pp. 832–839 (2000).
[2] 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).
[3] 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).

[4] Y. Kaneda and J. Ohga, “Adaptive microphone-array system for noise reduction,” *IEEE Trans. Acoust. Speech Signal Process.*, **ASSP-34**, 1391–1400 (1986).

[5] L. J. Griffiths and C. W. Jim, “An alternative approach to linearly constrained adaptive beamforming,” *IEEE Trans. Antennas Propag.*, **AP-23**, 27–34 (1982).

[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] 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).

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