Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Acoustic Array Systems: Paper ICA2016-686

Advanced delay-and-sum beamformer with deep neural network

Mitsunori Mizumachi [(a)] **, Maya Origuchi** [(a)]

[(a)] Kyushu Institute of Technology, Japan, mizumach@ecs.kyutech.ac.jp

Abstract

Signal enhancement can be achieved by spatial filtering with multiple microphones, that is, a microphone array. A traditional delay-and-sum beamformer enhances the target signal without spectral distortion, but could not control its beam-pattern, except for the arrival direction of the target signal. The beam-patten is flexibly designed by introducing channel-dependent weights or filters in delay-and-sum beamforming. The channel weights can be optimized using a neural network in order to achieve superdirectivity. In this study, an advanced delay-and-sum beamformer is proposed to make the main lobe much sharper and decrease grating lobes. The advanced beamformer substitutes a deep neural network for the conventional three-layered neural network. The proposed method also adopts a non-equally-spaced microphone arrangement and sub-band beamforming to prevent spatial aliasing. It is confirmed that the proposed beamformer has significant advantages both in sharpening the main-lobe and getting rid of grating lobes over conventional beamformers.

Keywords: Delay-and-sum beamformer, channel-dependent weight, deep neural network, non-equally-spaced microphone arrangement

Advanced delay-and-sum beamformer with deep neural network

1 Introduction

Beamforming has been one of important issues in acoustical signal processing [1]. It enables to detect and enhance acoustic signals in adverse conditions. A wide variety of beamformers have proposed for several decades. The traditional beamformers have been designed analytically and adaptively. In the field of information processing, a machine learning technique with huge amount of training data is a representative approach in non-linear optimization problems. A neural network can be an alternative approach to optimizing the beamformers. Kobatake *et al.* proposed a pioneering superdirective beamformer with a three-layered neural network structure [2]. The neural-network-based beamformers became a popular for the narrow-band antenna applications [3-5]. It is, however, difficult to deal with wide-band acoustical signals, although non-linear beamformers with the learning schemes based on neural networks have been investigated for acoustical applications [6-8].

In this paper, an advanced delay-and-sum beamformer is proposed with a deep neural network and an optimized microphone arrangement. It is expected to further achieve superdirectivity by using a four-layered neural network instead of the conventional three-layered neural network. There is another annoying problem of the appearance of grating lobes, that is, spatial aliasing in beamforming. The spatial aliasing is caused by spatial sampling using a microphone array, and is theoretically explained with the wavelength of the signal and the spacing of the neighboring microphones. The proposed beamformer employs a carefully designed microphone array, of which microphone spacing is optimized and three nesting sub-arrays achieve sub-band beamforming. In optimizing the proposed beamformer, it is also important to prepare suitable training data. In this paper, a single sinusoidal signal and wide-band random noises are prepared as the training data obtained by the microphone array. Feasibility of the proposed beamformer is confirmed by computer simulation.

2 Delay-and-sum beamformer

Delay-and-sum beamforming is a traditional means of spatial filtering. It can be achieved by linear signal processing, where multiple observed signals are phase-adjusted and added, and can be simply implemented both in analogue and digital manners. A target signal coming from a desired direction is not distorted by delay-and-sum beamforming. Those advantages have made delay-and-sum beamformers popular in acoustical signal processing. On the other hand, a delay-and-sum beamformer is not superior in efficiency to state-of-the-art beamformers. A delay-and-sum beamformer needs a number of microphones to form a sharp main love, especially in the low frequency range. It controls only the main lobe in the directivity pattern, and does not turn attention to the directions except the look direction. It means that the target signal is emphasized by synchronous addition and signals coming from the undesired directions are weaken by phase interference.

A filter-and-sum beamformer introduces a FIR filter into the phase-adjustment stage of delay-and-sum beamforming. The filter-and-sum beamforming is also linear in the viewpoint of signal processing, although it allows beamformers to control side lobes. Kobatake *et al.* proposed a new framework of non-linear beamforming, where a three-layered neural network was employed to achieve superdirectivity. The non-linear beamforming yielded the distortion on the target signal, but its superdirectivity is more advantageous compared with its defects.

3 Advanced delay and sum beamformer

3.1 Overview

An advanced delay-and-sum beamformer is proposed in order to achieve further superdirectivity and reduce grating lobes caused by spatial aliasing. The grating lobes appear in the specific frequencies, which can be theoretically determined due to the relationship between the wavelength of the signal and the spacing of the neighboring microphones. This problem can be solved by carefully designing the microphone arrangement. The proposed beamformer also aims at sharpening the main lobe using a deep neural network, which substitutes for the conventional three-layered neural network. The deep neural network optimizes the channel-dependent weights in the delay and sum beamformer.

3.2 Optimization of microphone arrangement

A microphone array must be carefully designed to prevent spatial aliasing in the target frequency range. It is well known that the sub-band beamforming with the sub-arrays, which have different microphone spacing for each sub-array, solves the problem on spatial aliasing. A nesting arrangement of microphones is a popular solution aiming at reducing of the array size and making the best use of microphones. Flanagan *et al.* designed a large-scale microphone array, where microphones are orderly arranged at the interval of the octave scale, that is, "power of two" [9].

The authors have also proposed the efficient nesting microphone arrangement [10]. It is found that the microphone arrangement according to the "power of three" rule is superior in signal enhancement to the octave arrangement. The "power of three" arrangement, however, requires a wider microphone array compared to the octave array. Therefore, the proposed beamformer employs the "multiple of three" rule as shown Fig. 1.

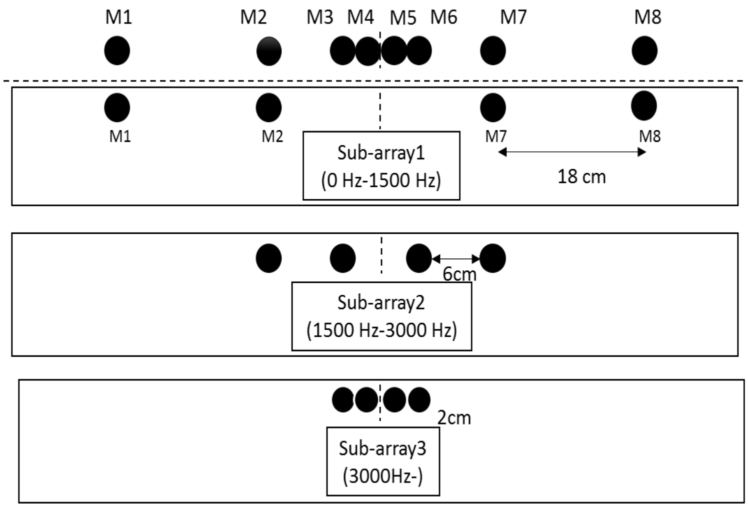

Figure 1: 8-ch microphone array with "multiple of three" arrangement

The proposed microphone array consists of three sub-arrays: Sub-array 1 with the microphone spacing of 18 cm up to 1.5 kHz, Sub-array 2 with the spacing of 6 cm from 1.5 kHz to 3 kHz, and Sub-array 3 with the spacing of 2 cm over 3 kHz. Microphones, M1 and M8, are solely used for the Sub-array 1, and microphones, M4 and M5, are used only for the Sub-array 3. Microphones, M2 and M7, are mutually used for the Sub-array 1 and Sub-array 2, and microphones, M3 and M6, are also mutually used for the Sub-array 2 and Sub-array 3.

3.3 Optimization of channel-dependent weight

The proposed beamformer prepares the channel-dependent weight for delay-and-sum beamforming, and the weights are optimized using a deep neural network. A preliminary experiment with various depths of multiple-layered neural networks suggests that a conventional three-layered neural network is not sufficient to optimize the channel-dependent weights for the 8-ch non-equally-spaced microphone array. Assuming that enough amount of training data are prepared, a deep neural network with more layers achieves superdirectivity, but it might cause overfitting to the training data and requires immense costs for training. Therefore, a four-layered neural network is employed in optimizing the channel weights.

The proposed beamformer trains the channel-dependent weights for each sub-array by supervised learning. A set of wide-band random noises is prepared as training data for the observations of the microphone arrays, where the signals come from 0 degree to 90 degrees at the interval of 10 degrees. The supervised signals are the training data itself and zero sequences for the input data corresponding to desired direction and undesired directions, respectively. The back-propagation algorithm is used for training the neural network, where initial channel-dependent weights are randomly set in [-1, 1].

4 Performance evaluation

Performance of the proposed beamformer is evaluated by computer simulation. In order to investigate the feasibility of training data, a single sinusoidal signal and the wide-band random noises are prepared for training the proposed beamformer, because the referenced conventional non-linear beamformer with a neural network [2] was trained with the sinusoidal signal. The traditional delay-and-sum beamformer, which does not employ a neural network, is also prepared as a reference.

Figure 1 and 2 illustrate beam-patterns for the proposed beamformer trained with the wide-band random noises in solid lines, conventional method trained with the single sinusoidal signal in broken lines, and the traditional delay-and-sum beamformer in dotted lines, respectively.

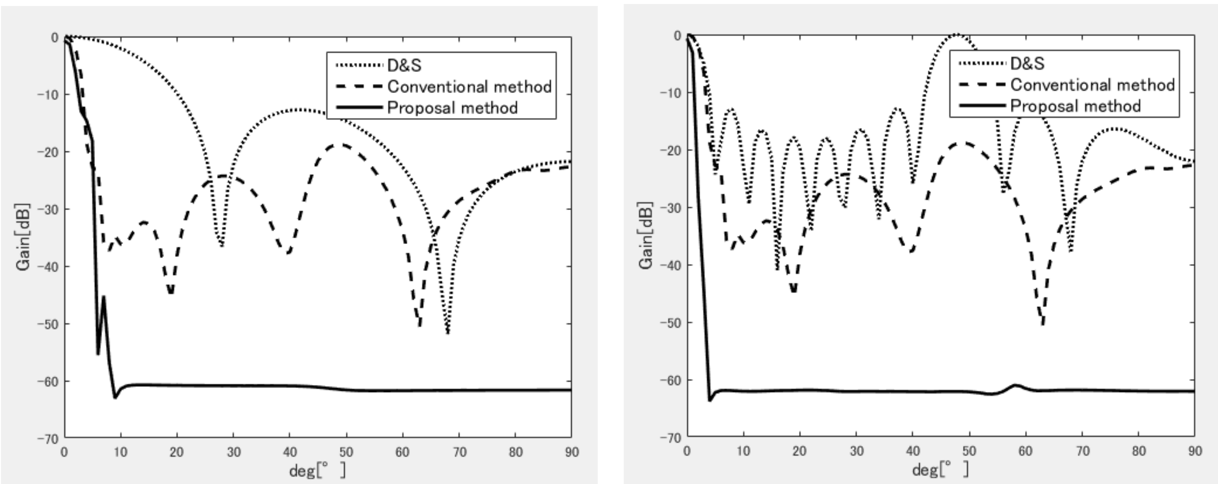

Figure 2: Beam-patterns in 1 kHz (left panel) and 5 kHz (right panel) for conventional delay-and-sum beamformer (dotted line), conventional beamformer trained with single sinusoidal signal (broken line), and proposed beamformer trained with wide-band random noise (solid line).

The proposed beamformer succeeds in achieving superdirectivity without any grating lobes in both 1 kHz and 5 kHz. It is obvious that the proposed beamformer has a significant advantage over the conventional method and the traditional delay-and-sum beamformer.

5 Conclusions

In this paper, an advanced delay-and-sum beamformer is proposed to simultaneously achieve superdirectivity and prevent the occurrence of grating lobes. Channel-dependent weights are introduced for delay-and-sum beamforming, and are optimized by a deep neural network. The four-layered neural network with the informative training data attains to sharpen the main lobe in the beam-pattern at the desired direction, and the non-equally-spaced microphone arrangement contributes to make the grating lobes disappear. Computer simulations have verified the feasibility of the proposed method compared with conventional linear and non-linear beamformers. Future works include the performance evaluation of the proposed beamformer in signal enhancement under real environments.

References

[1] Brandstein, M. and Darren W. (eds.). *Microphone arrays: signal processing techniques and applications*. Springer, 2013.

[2] Kobatake, H.; Morita W.; Yano, Y. Super directive sensor array with neural network structure, *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Vol. 2, 1992, pp. 321-324.

[3] Chang, P-R; Yang, W-H; and Chan, K-K. A neural network approach to MVDR beamforming problem. *IEEE Trans. Ant. and Propag.*, vol. 40, no. 3, 1992, pp. 313-322.

[4] Southall, H. L.; Simmers, J. A.; O'Donnell, T. H. Direction finding in phased arrays with a neural network beamformer. *IEEE Trans. Ant. and Propag.*, vol. 43, no. 12, 1995, pp. 1369-1374.

[5] Zooghby, A. H. E.; Christodoulou, C. G.; Georgiopoulos, M. Neural network-based adaptive beamforming for one- and two-dimensional antenna arrays. *IEEE Trans. Ant. and Propag.*, vol. 46, no. 12, 1998, pp. 1891-1893.

[6] Dahl, M.; Claesson, I. A neural network trained microphone array system for noise reduction. *Proc. IEEE Signal Processing Society Workshop*, 1996, pp. 311-319.

[7] Song, X.; Wang, J.; Han, Y.; Tian, D. Neural Network-Based Robust Adaptive Beamforming. *Proc. International Joint Conference on Neural Network Proceedings*, 2006, pp. 1758-1763.

[8] Iseki, A.; Ozawa, K.; Kinoshita, Y. Neural network-based microphone array learning of temporal-spatial patterns of input signals. *Proc. IEEE 3rd Global Conference on Consumer Electronics*, 2014, pp. 88-89.

[9] Flanagan, J. L.; Berkley, D. A.; Elko, G. W; West, J. E.; Sondhi, M. M. Autodirective Microphone Systems, *Acta Acustica*, Vol. 73, No. 2, 1991, pp. 58-71.

[10] Hayashi, H.; M. Mizumachi, M. Speech enhancement by non-linear beamforming tolerant to misalignment of target source direction, *Journal of the Institute of Industrial Applications Engineers*, Vol.1, No.2, 2013, pp. 97-104.