24th INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
23–27 July 2017, London

NON-LINEAR BROADBAND BEAMFORMER WITH DEEP NEURAL NETWORK

Mitsunori Mizumachi

Kyushu Institute of Technology, Faculty of Engineering, Kitakyushu, Fukuoka, Japan.
email: mizumach@ecs.kyutech.ac.jp

Maya Origuchi, Yuya Nishijima

Kyushu Institute of Technology, Graduate School of Engineering, Kitakyushu, Fukuoka, Japan.

Speech enhancement is indispensable for human-to-machine speech interfaces, because speech signals are extremely fragile against acoustical interferences. Spatial filtering, that is, beamforming, is one of the popular solutions using multiple microphones. In the case of linear beamformers, a number of microphones are required to make the main lobe sharp at the desired direction. On the other hand, non-linear beamformers efficiently form the sharp main lobe compared with the linear beamformers. Kobatake et al. proposed a pioneering non-linear beamformer with a neural network. This method considered the narrowband optimization of the network structure, and succeeded in both achieving a sharper main lobe and decreasing grating lobes. Narrowband non-linear beamformers have been steadily developed in the field of antenna and propagation. In this paper, a non-linear broadband beamformer with a deep neural network is proposed for speech enhancement. The proposed beamformer substitute a deep neural network for a conventional neural network, and non-equally-spaced microphone arrangement is employed for making grating lobes disappear. It is confirmed that the proposed beamformer can achieve ideal beam-pattern under a real environment.

Keywords: microphone array, beamforming, deep neural network, speech enhancement

1. Introduction

Beamforming has been one of important issues in acoustical signal processing [1] as well as radar and radio applications. It enables to detect and enhance target signals in adverse conditions. A wide variety of beamformers have proposed for several decades. The traditional beamformers have been analytically and adaptively designed such as a delay-and-sum beamformer [2] and the AMROR [3]. In the field of information processing, machine learning 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 [4]. Neural network-based beamformers became popular for narrow-band antenna applications [5-7]. It is, however, difficult for those beamformers to deal with wide-band acoustical signals, although various non-linear beamformers with the learning schemes based on neural networks have been investigated for acoustical applications [8-10].

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.

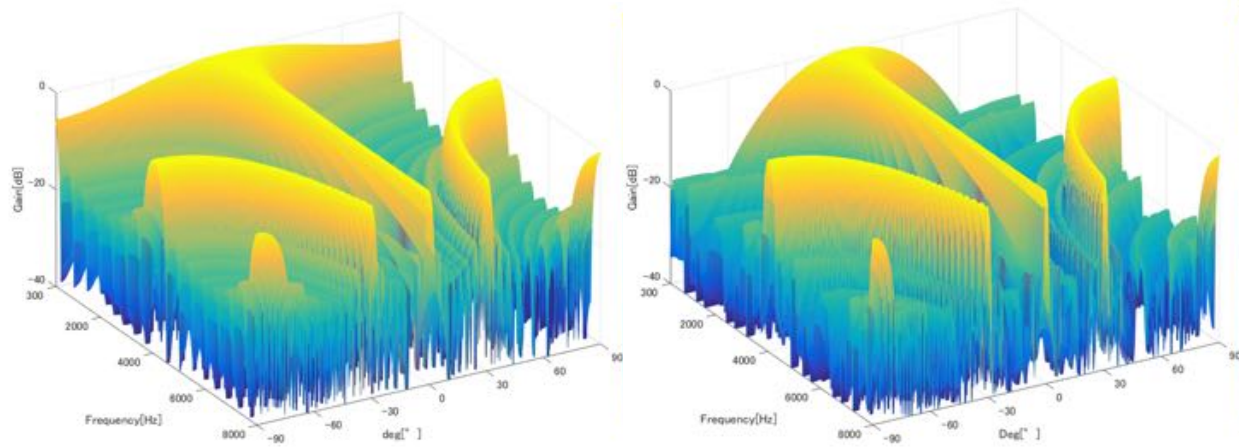

Figure 1: Beam-patterns in spatial and frequency domains for delay-and-sum beamformer (left panel) and non-linear beamformer with neural network (right panel).

The spatial aliasing is caused by spatial sampling using a microphone array, and is theoretically explained based on the wavelength of the signal and the spacing of neighbouring 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, wide-band random noises and diffused noises are used for training the beamformer. Feasibility of the proposed beamformer is confirmed under a real environment.

2. Conventional beamformers

The delay-and-sum beamforming is a traditional means of spatial filtering [1, 2]. It can be achieved by linear signal processing, where multiple observed signals are phase-adjusted and summed up, 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 novel 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.

Figure 1 shows the beam-patterns in spatial and frequency domains for the delay-and-sum beamformer and the non-linear beamformer with a neural network, respectively. Eight microphones are equally-spaced with the spacing of 0.084 m, and the target signal comes from 0 degree, that is, the front of the array. In Fig. 1, the non-linear beamformer is superior in the sharpness of the main lobe to the delay-and-sum beamformer, although both beamformers have grating lobes in high frequency due to spatial aliasing.

3. Proposed non-linear broadband beamformer

3.1 Overview

A non-linear broadband beamformer is proposed in order to achieve further superdirectivity and reduce grating lobes. 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 neighbouring 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.

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 well-known 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" [2].

The authors have also proposed the efficient nesting microphone arrangement [11]. 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. The proposed microphone array consists of three nesting sub-arrays: Sub-array 1 with the microphone spacing of 15.75 cm up to 1.5 kHz, Sub-array 2 with the spacing of 5.25 cm from 1.5 kHz to 3 kHz, and Sub-array 3 with the spacing of 1.75 cm over 3 kHz. Microphones.

3.3 Optimization of channel weight using deep neural network

The proposed beamformer prepares the channel-dependent weight for non-linear 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, which is described in Section **3.2**. Assuming that enough amount of training data can be 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 and diffused noises are 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 signals coming from the 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].

Table 1: Data set for training the proposed non-linear beamformer

	(directional random noise) : (diffused random noise)
Training data set 1	100:0
Training data set 2	90:10, 100:0
Training data set 3	80:20, 90:10, 100:0
Training data set 4	70:30, 80:20, 90:10, 100:0

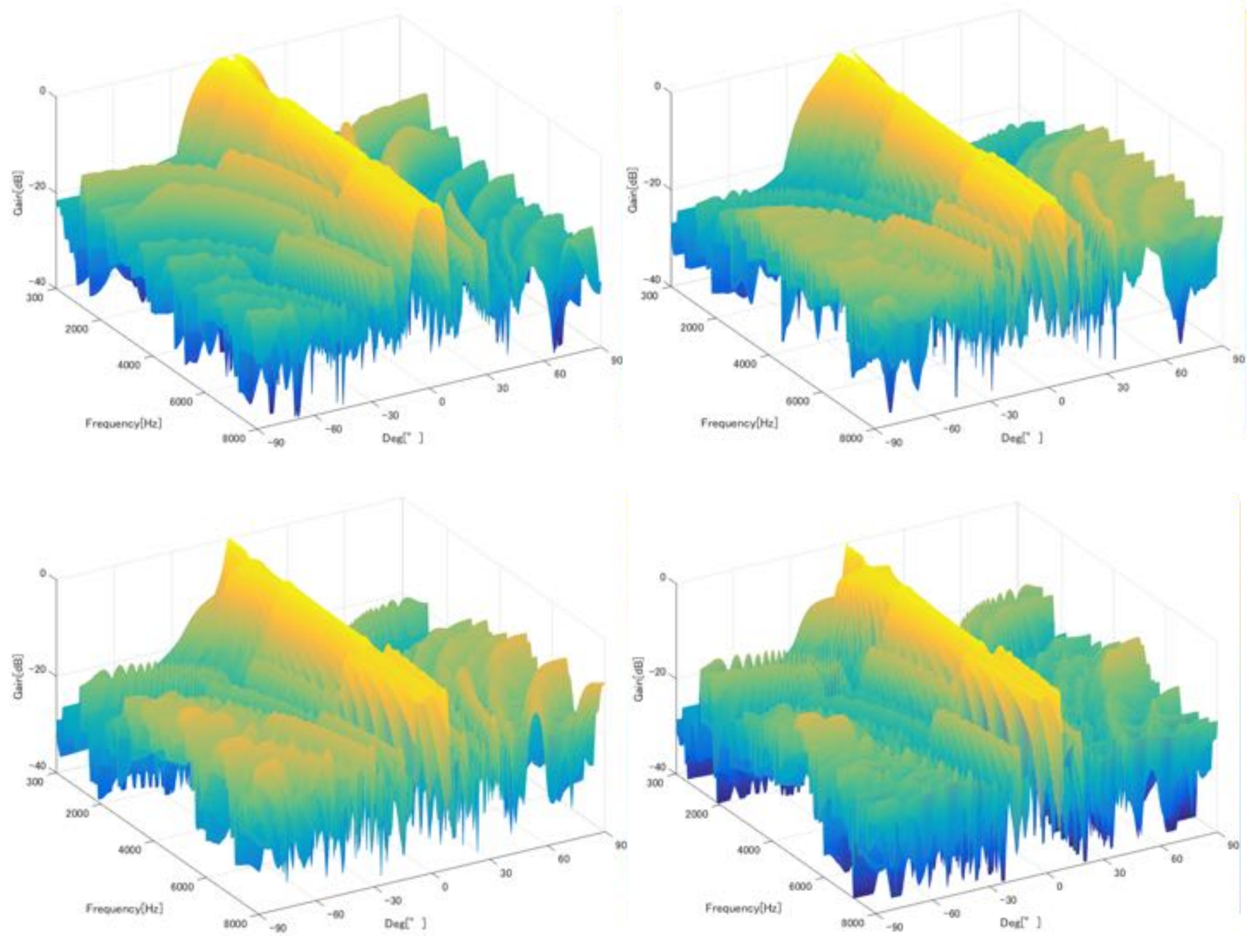

Figure 2: Beam-patterns of the proposed beamformers with the training data set 1 (upper left panel), training data set 2 (upper right panel), training data set 3 (lower left panel), and training data set 4 (lower right panel).

4. Performance evaluation

Performance of the proposed beamformer is evaluated under a soundproofed room. Both directional wide-band random noises in [300 Hz, 24,000 Hz] and diffused versions of those noises are prepared as the training data, although the referenced conventional non-linear beamformer with the three-layered neural network [4] was trained with the sinusoidal signal. Four training data sets are prepared as shown in Table 1. Resultant beam-patterns obtained with the different training data sets are shown in Fig. 2. It is confirmed that the beam-patterns vary depending on the training data.

Feasibility of the proposed beamformer is also investigated in the viewpoint of noise reduction. Signal-to-noise ratio and spectral distortion in [300 Hz, 3,400 Hz] are employed as objective distortion measures.

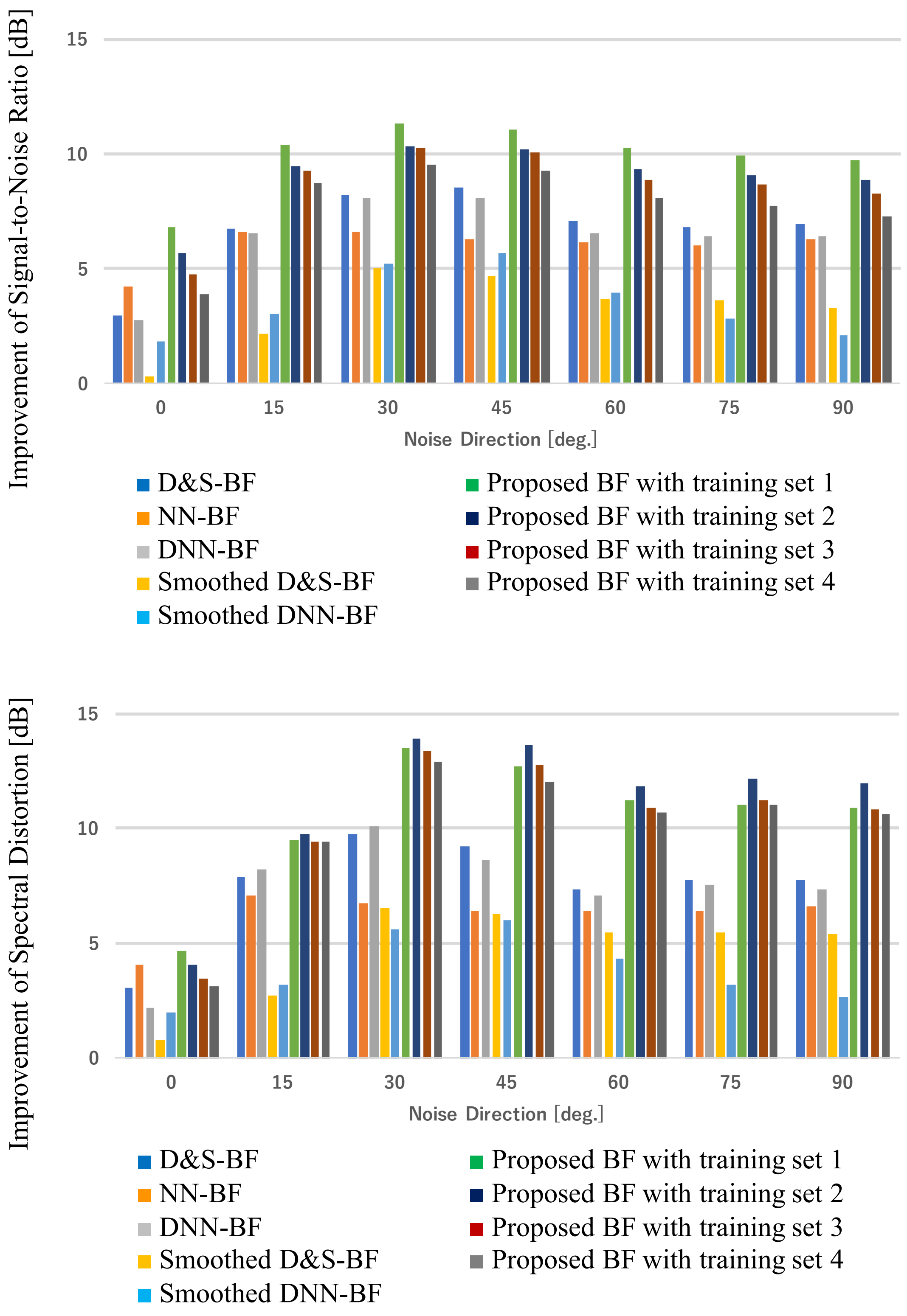

Figure 3: Improvements of signal-to-noise ratio and spectral distortion by spatial filtering with various beamformers in upper and lower panels, respectively.

Figure 3 gives the improvements of signal-to-noise ratio and spectral distortion by spatial filtering with various beamformers: Delay-and-sum beamformers with equally-spaced and non-equally-spaced microphone arrangements (D&S-BF and Smoothed D&S-BF), non-linear beamformers with the equally-spaced microphones trained by three-layered and four-layered neural networks using sinusoid signals (NN-BF and DNN-BF), non-linear beamformers with the equally-spaced microphones trained by four-layered neural networks with sinusoid signals (Smoothed DNN-BF), and the proposed beamformers with different training sets. The proposed beamformer is superior to conventional beamformers, and the suitable training data set is different according to the distortion measures. It is necessary to further investigate the relation between training data and noise reduction performance.

5. Conclusions

A non-linear broadband beamformer is proposed to simultaneously achieve superdirectivity and prevent the occurrence of grating lobes. The channel-dependent weights in the four-layered neural network are optimized with directional and diffused wide-band random noises. The non-equally-spaced microphone arrangement contributes to make the grating lobes disappear. An experiment under a real environment have verified the feasibility of the proposed method compared with conventional linear and non-linear beamformers trained with sinusoidal signals. Future works include the performance evaluation of the proposed beamformer in complex acoustic scenes with reverberation.

Acknowledgment

This work was supported by KAKENHI 16H01734.

REFERENCES

1 Brandstein, M. and Darren W. (eds.). Microphone arrays: signal processing techniques and applications. Springer, 2013.
2 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.
3 Kaneda, Y; Ohga, J. Adaptive microphone-array system for noise reduction, IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. 34, Issue 6, 1986, pp. 1391 – 1400.
4 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.
5 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.
6 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.
7 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.
8 Dahl, M.; Claesson, I. A neural network trained microphone array system for noise reduction. *Proc. IEEE Signal Processing Society Workshop*, 1996, pp. 311-319.
9 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.
10 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.
11 Hayashi, H.; 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.