Deep Neural Network for Source Localization Using Underwater Horizontal Circular Array

Zhaoqiong Huang[1,2], Ji Xu[1,2], Chen Li[1,2], Zaixiao Gong[2,3], Jielin Pan[1,2] and Yonghong Yan[1,2]
[1] *Key Laboratory of Speech Acoustics and Content Understanding, Institute of Acoustics, Chinese Academy of Sciences*
[2] *University of Chinese Academy of Sciences*
[3] *State Key Laboratory of Acoustics, Institute of Acoustics, Chinese Academy of Sciences*
Beijing, China
huangzhaoqiong@hccl.ioa.ac.cn,xuji@hccl.ioa.ac.cn

***Abstract*—This paper applies deep neural network (DNN) to source localization in a shallow water environment using underwater horizontal circular array. The proposed method can discriminate source locations in a three-dimension space. The proposed method adopts a two-stage scheme, incorporating feature extraction and DNN analysis. In feature extraction step, the eigenvectors corresponding to the modal signal space, which are shown to be able to represent the propagating modes of the sound source, are extracted as the input feature of DNN. The eigenvectors are obtained by applying eigenvalue decomposition (EVD) of the covariance matrix of the received multi-channel signal. In DNN analysis step, time delay neural network (TDNN) is used to construct the mapping relationship between the eigenvectors and the source locations, because it is capable of making use of sequential information of the source signal. The output of the network is the source location estimates. Several experiments are conducted to demonstrate the effectiveness of the proposed method.**

***Index Terms*—Deep neural network, source localization, horizontal circular array, modal signal space, shallow water environment.**

I. Introduction

Source localization in a shallow water environment has attracted a lot of attention in the past several decades. To reduce the dependence on environmental information, many data-based localization methods are presented that directly estimate source location from acoustic data. Passive source ranging with a guide source estimates the unknown source range by correlating the field from the guide source at one frequency component with the unknown source field at a different range, for a different frequency component [1]. The array/waveguide invariant is proposed for robust source-range estimation in [2], [3]. Besides, numerous machine leaning techniques have been introduced to source localization [4]– [8]. These methods take source localization as a classification or regression task, commonly based on conventional classifiers or shallow feed-forward neural networks (FNNs) [5]– [8].

Deep learning is a famous data-driven technique, which achieves significant results on image recognition, speech recognition, natural language processing etc. in recent years for its powerful modeling capability. In our previous study [9], we firstly applies DNN to source localization task and demonstrates the effectiveness of DNN based method using linear array. However, linear arrays can only distinguish sources in a two-dimension space. In this paper, we extend DNN based method to source localization using underwater horizontal circular array (HCA), which can discriminate the sources in a three-dimension space.

The proposed method adopts a two-stage scheme including feature extraction and DNN analysis. The eigenvectors corresponding to the modal signal space, which are taken as the input feature of the neural network, are decomposed from the covariance matrices of the data field at different frequencies. Then, the estimates of azimuth, range, and depth are given by regression network learning. Since the time delay neural network (TDNN) [10] has the advantage of modeling the temporal dynamics in sound signal, it is chosen as the basic network architecture. After DNN training, the unknown source location is obtained by feeding the extracted feature into the trained model.

The rest of the paper is organized as follows. Section II describes the proposed method. Section III gives various experiments for evaluation, and Section IV concludes this work.

II. Proposed Method

The DNN based source localization system is illustrated in Fig. 1. In the training stage, the feature extraction module aims at extracting features that can represent source location information effectively. The eigenvectors associated with the

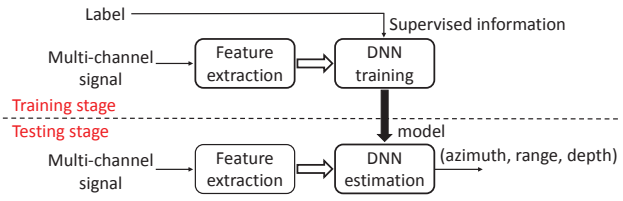

Fig. 1. The overview of the source localization system.

This work is partially supported by the National Natural Science Foundation of China (Nos. 11590770-4) and the Innovation Foundation of Chinese Academy of Sciences (No. CXQZ201701).

978-1-5386-1654-3/18/$31.00 ©2018 IEEE

modal signal space are extracted as the input feature for DNN by eigenanalysis. Then DNN learns the regressive transform from the eigenvectors to source location in the supervised mode. In the testing stage, the unlabeled test data are feed into the trained model, then the location estimates are given as continuous values.

A. Feature extraction

Let's consider a single wide-band sound source impinges on a HCA of K sensors in a far-field scenario. Using the matrix notation, the pressure field [11] received by the sensors is described as

$$\boldsymbol{P} = a\boldsymbol{H}\boldsymbol{S} + \boldsymbol{N}, \tag{1}$$

where a denotes the complex Gaussian random amplitude of the source, $\boldsymbol{S} \in \mathbb{C}^M$ with $S_m(r,z) = (2\pi/k_m r)^{1/2}\Psi_m(z)e^{jk_m r}$, M ($M < K$) denotes the mode number in the water column (higher modes are treated as noise), k_m^2 is the eigenvalue associated with the m-th mode, $\boldsymbol{N} \in \mathbb{C}^K$ denotes the additive noise, and $\boldsymbol{H} \in \mathbb{C}^{K\times M}$ with $H_{m,k} = \Psi_m(z_k)$. $\boldsymbol{P} = [P_1, \ldots, P_K]^T \in \mathbb{C}^K$, where $(\cdot)^T$ denotes the transpose operation and the pressure field at depth z_k due to a source at (r,z) is represented as

$$P_k = b\sum_{m=1}^{M} H_{m,k}S_m(r,z), \tag{2}$$

where $b = \frac{j}{\rho(z_s)\sqrt{8\pi}}e^{-j\pi/4}$, k_m^2 is the eigenvalue associated with the m-th mode, $\rho(z_s)$ denotes the density at the source, r is the horizontal distance between the source and the k-th hydrophone, and z is the source depth.

The covariance matrix at a single frequency over D snapshots is expressed as

$$\boldsymbol{R}(f) = \frac{1}{D}\sum_{d=1}^{D}\boldsymbol{P}_d(f)\boldsymbol{P}_d^{+}(f), \tag{3}$$

where f denotes the frequency, $(\cdot)^+$ denotes the Hermitian transpose. Applying eigenvalue decomposition (EVD) to $\boldsymbol{R}(f)$,

$$\begin{aligned}\boldsymbol{R}(f) &= \boldsymbol{\Lambda}_f\boldsymbol{\Sigma}_f\boldsymbol{\Lambda}_f^{+} \\ &= \boldsymbol{\Lambda}_f^S\boldsymbol{\Sigma}_f^S\boldsymbol{\Lambda}_f^{S+} + \boldsymbol{\Lambda}_f^N\boldsymbol{\Sigma}_f^N\boldsymbol{\Lambda}_f^{N+},\end{aligned} \tag{4}$$

where the eigenvectors and eigenvalues are obtained as $\boldsymbol{\Lambda}_f = [\mathbf{e}_{f,1}, \ldots, \mathbf{e}_{f,K}] \in \mathbb{C}^{K\times K}$ and $\boldsymbol{\Sigma}_f = diag[\lambda_{f,1}, \ldots, \lambda_{f,K}]$, where the eigenvalues are sorted in descending order. $\boldsymbol{\Sigma}_f^S$ and $\boldsymbol{\Lambda}_f^S = [\mathbf{e}_{f_i,1}, \ldots, \mathbf{e}_{f_i,M'}] \in \mathbb{C}^{K\times M'}$ are eigenvalues and eigenvectors corresponding to the modal signal space and $\boldsymbol{\Sigma}_f^N$ and $\boldsymbol{\Lambda}_f^N$ corresponds to the noise space.

The eigenvectors with relative larger eigenvalues ($\boldsymbol{\Lambda}_f^S$), which are considered to be the main feature of the propagating modes of an assumed source location, are used as the input feature of DNN. The remaining eigenvectors are disregarded to suppress the noise. The block diagram of the proposed method is shown in Fig. 2. As the eigenvectors obtained by EVD are complex values, they can not be directly addressed by DNN. The real and imaginary part of the eigenvectors are concatenated as the input vector,

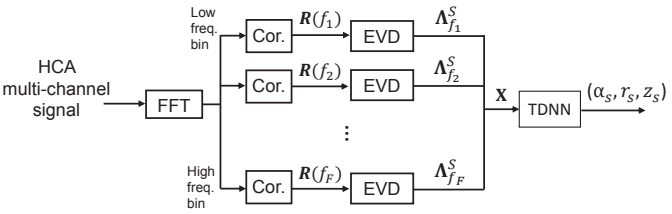

Fig. 2. The block diagram of the proposed method.

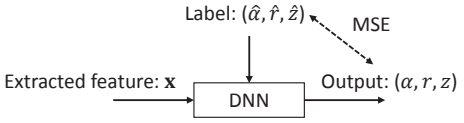

Fig. 3. Block diagram of DNN learning.

$$\mathbf{x} \triangleq \bigcup_{m}\bigcup_{\iota}\Big[\mathcal{R}(\boldsymbol{\Lambda}_{f_\iota,m}), \Im(\boldsymbol{\Lambda}_{f_\iota,m})\Big], \iota = 1, \ldots, F, d = 1, \ldots, M', \tag{5}$$

where ι denotes the frequency index, F denotes the number of frequency bins. In this paper, the source location is represented by a three-dimension vector including azimuth (α_s) in degree, range (r_s) in kilometer, and depth (z_s) in meter. Note that r_s, which is different from r, is the horizonal distance between the source and array center.

B. DNN analysis

Because DNN has the strong capability of modeling the complex non-linear relationship, it is used to construct the regression model between the eigenvector and source location, as illustrated in Fig. 3,

$$(\alpha, r, z) = \mathcal{F}(\mathbf{x}), \tag{6}$$

where $\mathbf{x}$ denotes the feature and $\mathcal{F}(.)$ denotes the regressive transform function obtained by DNN training.

The network parameters are updated by minimizing the mean square error (MSE) between the estimated source location and the reference source location, given by

$$E = \frac{1}{L}\sum_{l=1}^{L}\Big[(\alpha_l - \alpha_l^{'})^2 + (r_l - r_l^{'})^2 + (z_l - z_l^{'})^2\Big], \tag{7}$$

where $(\alpha_l^{'}, r_l^{'}, z_l^{'})$ and (α_l, r_l, z_l) denote the reference and estimated values respectively, and L denotes the sample number. The back-propagation (BP) [17] algorithm with stochastic gradient descent (SGD) [18] is used to optimize the network parameters.

III. Experiments

A. Simulation setup

1) Acoustic environmental model: To investigate the performance of the proposed method, we conducted several experiments in the 80 m ocean environment, where a 50-element uniform HCA was deployed at the bottom of the

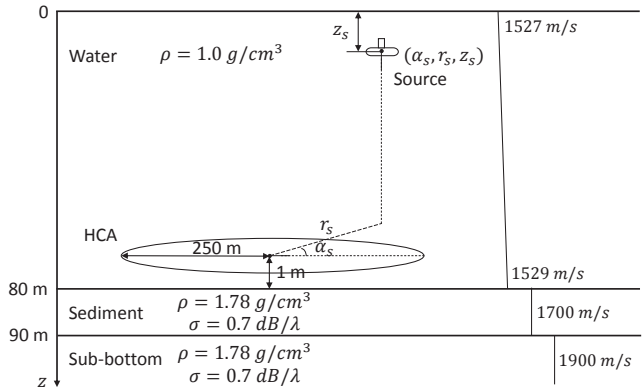

Fig. 4. Schematic diagram of the simulated acoustic environmental model.

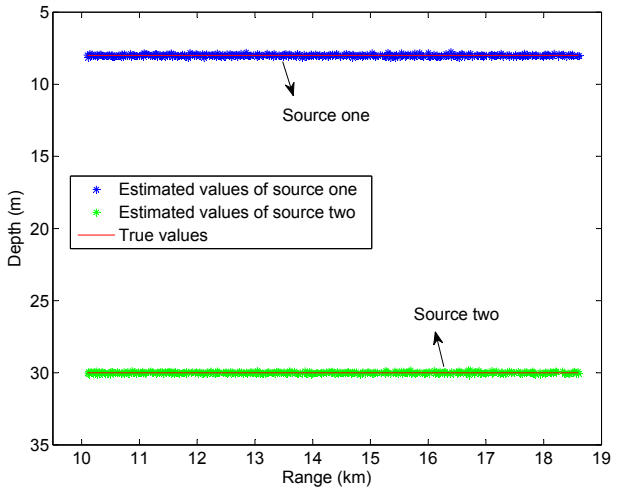

Fig. 5. Depth estimates versus range estimates of test source.

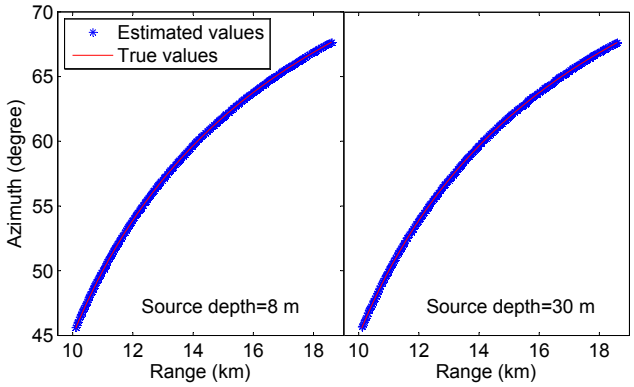

Fig. 6. Azimuth estimates versus range estimates of test source.

water. The schematic diagram of the simulated environment was illustrated in Fig. 4. The radius of HCA was 250 m. The depths of water and the sediment layer were 80 m and 10 m. The sound speed profile (SSP) was also plotted in Fig. 4. ρ and σ in Fig. 4 denote the density and the attenuation coefficient.

2) Data description: Acoustic data was simulated using KRAKEN. There were totally 53000 training samples and 5300 test samples. The training data and test data were mutually different. The depth of training or test sources were between 3.5 m and 35 m. For each source, the depth was kept fixed, while the range was from 10 km to 18.5 km and azimuth from 45° to 67°. The signal-to-noise ratio (SNR) was set to about -20 dB.

3) Parameters for feature extraction: The signal transformed to frequency domain by operating fast Fourier transformation (FFT). The frame length was about 5.5 s. The bandwidth used for feature extraction was set to $[100, 300]$ Hz. Ten eigenvectors at sixteen frequency bins were extracted as the input feature, thus the feature per frame included 16000 $(50 \times 10 \times 16 \times 2)$ dimensions.

4) Parameters for TDNN: The configuration of TDNN was that of 8 layers (one input layer + six hidden layers + one output layer) with 1024 hidden nodes. The units were spliced from $t-1$ to $t+1$ at the input layer and the second hidden layer [19]. No frame was spliced at the first, and third to sixth hidden layers. Current output was determined by the inputs from $t-2$ to $t+2$ (totally 5 frames) in terms of the whole framework. The rectified linear units (ReLU), $f(x) = \max(0, x)$, was used as the activation function [20]. The parameters of the network were optimized by the BP algorithm with SGD by taking MSE criterion as the cost function. The Kaldi [21] toolkit was utilized for DNN training. The initial learning rate was 0.001 and the batch for SGD was 512.

TABLE I
MAE AND MRE OF SOURCE LOCATION ESTIMATES.

	MAE	MRE
Azimuth	0.05	0.9%
Depth	0.05	0.9%
Range	0.02	0.2%

B. Results

The objective evaluation metrics used were the mean absolute error (MAE) and the mean relative error (MRE),

$$MAE = \frac{1}{Q}\sum_{q=1}^{Q}|x_q - x_q^{'}|, \tag{8}$$

$$MRE = \frac{1}{Q}\sum_{q=1}^{Q}\left|\frac{x_q - x_q^{'}}{x_q^{'}}\right| \times 100\%, \tag{9}$$

where x represents the estimation value and $x^{'}$ represents the reference value. Q is the sample number. The MAE and MRE of range, depth, and azimuth averaged over all test samples are summarized in Table I. We can find that the MAE and MRE of azimuth, depth, and range consistently reach small values under a low SNR.

Besides, the estimated values of test source at depth 8 m and 30 m are plotted as examples to display the accuracy of azimuth, range, and depth estimates. The depth estimates versus range estimates is shown in Fig. 5. From Fig. 5, we can see that the proposed method can give reliable estimates to source's trajectory generally. Also, the azimuth estimates versus range estimates and the range estimates versus true ranges are respectively shown in Fig. 6 and Fig. 7. The results

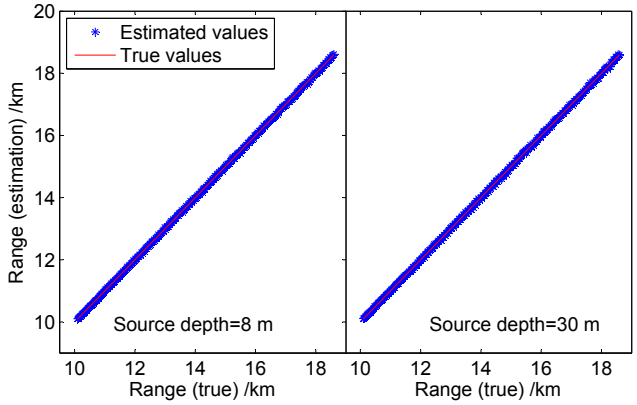

Fig. 7. Estimated ranges versus true ranges of test source.

show that the proposed method can give accurate range and azimuth estimates. The experimental results demonstrate DNN is effective for source localization task using underwater HCA.

IV. Conclusions

This paper applied DNN based source localization method to HCA scenario. The eigenvectors corresponding to modal signal space is extracted as the input feature. TDNN is adopted to construct the regressive transformation between the eigenvectors to source location. In contract to our precious works, it can discriminate sources in a three-dimension space. We will enhance the robustness of DNN based method under non-Gaussian noise environments in our future work.

References

[1] A. M. Thode, "Source ranging with minimal environmental information using a virtual receiver and waveguide invariant theory," *J. Acoust. Soc. Am.*, vol. 108, no. 4, 2000, pp. 1582–1594.

[2] C. Cho, H. C. Song, and W. S. Hodgkiss, "Robust source-range estimation using the array/waveguide invariant and a vertical array," *J. Acoust. Soc. Am.*, vol. 139, no. 1, 2016, pp. 63–69.

[3] H. C. Song and C. Cho, "Array invariant-based source localization in shallow water using a sparse vertical array," *J. Acoust. Soc. Am.*, vol. 141, 2017, pp. 183–188.

[4] R. Lefort, G. Real, and A. Drémeau, "Direct regressions for underwater acoustic source localization in fluctuating oceans," *Appl. Acoust.*, vol. 116, 2017, pp. 303–310.

[5] J. M. Ozard, P. Zakarauskas, and P. Ko, "An artificial neural network for range and depth discrimination in matched field processing," *J. Acoust. Soc. Am.*, vol. 90, no. 5, 1991, pp. 2658–2663.

[6] P. Zakarauskas, J. M. Ozard, and P. Brouwer, "Neural networks for independent range and depth discrimination in passive acoustic localization," *IEEE Transactions on Signal Processing*, vol. 41, no. 3, 1993, pp. 1394–1398.

[7] H. Niu, E. Reeves, and P. Gerstoft, "Source localization in an ocean waveguide using supervised machine learning," *J. Acoust. Soc. Am.*, vol. 142, no. 3, 2017, pp. 1176–1188.

[8] H. Niu, E. Ozanich, and P. Gerstoft, "Ship localization in Santa Barbara Channel using machine learning classifiers", *J. Acoust. Soc. Am.*, vol. 142, EL455-460, 2017.

[9] Z. Huang, J. Xu, Z. Gong, H. Wang, Y. Yan, "Source localization using deep neural networks in a shallow water environment," *J. Acoust. Soc. Am.*, 2018.

[10] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, 1989, pp. 328–339.

[11] C. L. Byrne, R. T. Brent, C. Feuillade, and D. R. DelBalzo, "A stable data-adaptive method for matched-field array processing in acoustic waveguides," *J. Acoust. Soc. Am.*, vol. 87, no. 6, 1990, pp. 2493–2502.

[12] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. M. Qian, P. Schwarz and et al., "The kaldi speech recognition toolkit," in *IEEE ASRU*, 2011.

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, 1986, pp. 533–536.

[14] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Computation*, vol. 18, no. 7, 2006, pp. 1527–1554.

[15] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, 2015, pp. 85–117.

[16] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang, "Phoneme recognition using time-delay neural networks," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 3, 1989, pp. 328–339.

[17] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, 1986, pp. 533–536.

[18] D. Povey, X. Zhang, and S. Khudanpur, "Parallel training of DNNs with natural gradient and parameter averaging," In *International Conference on Learning Representations: Workshop track*, 2015.

[19] V. Peddinti, D. Povey, S. Khudanpur, "A time delay neural network architecture for efficient modeling of long temporal contexts," *Interspeech 2015*, pp. 3214–3218, 2015.

[20] X. Glorot, A. Bordes , and Y. Bengio, "Deep Sparse Rectifier Neural Networks", in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, vol. 15, 2011, pp. 315–323.

[21] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. M. Qian, P. Schwarz and et al., "The kaldi speech recognition toolkit," in *IEEE ASRU*, 2011.