ICASSP 2018

A DEEP NEURAL NETWORK BASED METHOD OF SOURCE LOCALIZATION IN A SHALLOW WATER ENVIRONMENT

Zhaoqiong Huang, Ji Xu, Zaixiao Gong, Haibin Wang, and Yonghong Yan

Key Laboratory of Speech Acoustics and Content Understanding, Chinese Academy of Sciences

State Key Laboratory of Acoustics, Institute of Acoustics

Email: huangzhaoqiong@hccl.ioa.ac.cn

Abstract

This paper applies deep neural network (DNN) to source localization in a shallow water environment because of its powerful modeling capability and the little dependence on the prior knowledge of environmental parameters. The classical two-stage scheme is adopted, in which feature extraction and DNN analysis are independent steps. It firstly extracts the input feature from the observed signal received by underwater hydrophones. The eigenvectors associated with the modal signal space are decomposed from the covariance matrices of the data field at different frequencies, which are used as the input feature of DNN. The time delay neural network (TDNN) is exploited to model the long term feature representation and construct the regression model. The output is the source range-depth estimate. Several experiments using simulation and experimental data are conducted to evaluate the performance of the proposed method. The results demonstrate the effectiveness and potential of DNN for source localization. Particularly, experiments show that simulation data can be merged to train a general model for experimental data when lacking of sufficient training data in real-world environment.

Signal model

$$P(r_s, z_s, z_k) = ab\sum_{m=1}^{M}\frac{\Psi_m(z_s)\Psi_m(z_k)}{\sqrt{k_m r_s}}e^{jk_m r_s} + N(z_k),$$

a denotes the complex amplitude of the source, k_m^2 is the eigenvalue associated with the m-th mode, $\psi_m(z_s)$ and $\psi_m(z_k)$ denote the m-th mode eigenfunctions at the source and receiver, $N(z_k)$ denotes the additive noise at the k-th sensor, M $(M < K)$ denotes the mode number in the water column (higher modes are treated as noise).

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

Proposed Method

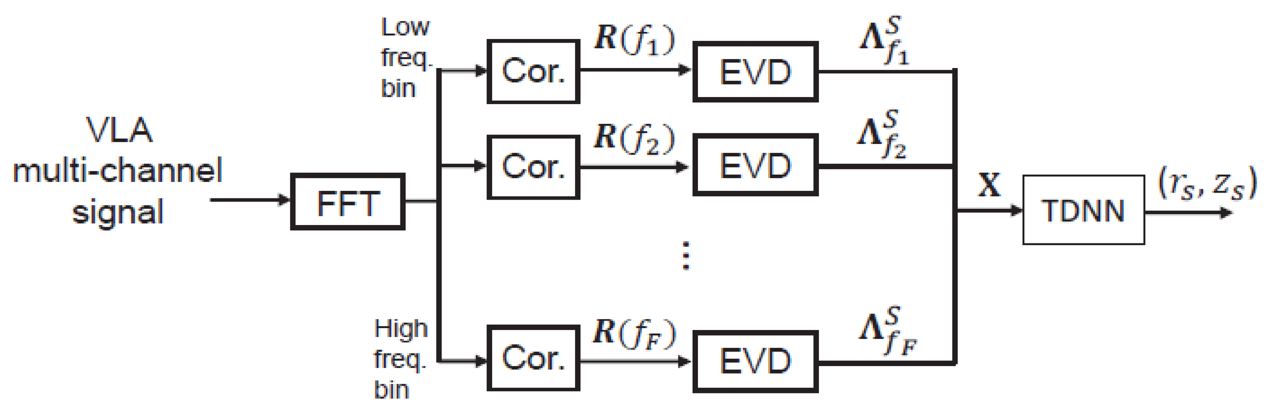

Fig. 1. Block diagram of the proposed method

1) Feature extraction

$$\boldsymbol{R}(f) = \frac{1}{D}\sum_{d=1}^{D}\boldsymbol{P}_d(f)\boldsymbol{P}_d^{+}(f), \quad = \boldsymbol{H}\boldsymbol{R}_S(f)\boldsymbol{H}^{+} + \boldsymbol{R}_N(f),$$

$$\boldsymbol{R}(f) = \boldsymbol{\Lambda}_f\boldsymbol{\Sigma}_f\boldsymbol{\Lambda}_f^{+} = \boldsymbol{\Lambda}_f^S\boldsymbol{\Sigma}_f^S\boldsymbol{\Lambda}_f^{S+} + \boldsymbol{\Lambda}_f^N\boldsymbol{\Sigma}_f^N\boldsymbol{\Lambda}_f^{N+},$$

The eigenvectors associated with the modal signal space is taken as the input feature of DNN.

$$\mathbf{x} \triangleq \bigcup_i \left[\mathcal{R}(\boldsymbol{\Lambda}_{f_i}^S), \Im(\boldsymbol{\Lambda}_{f_i}^S)\right], i = 1, \ldots, F.$$

2) DNN analysis

Cost function: MSE

BP algorithm with SGD

Learning rate: 0.001

Batch size: 512

$$E = \frac{1}{L}\sum_{l=1}^{L}\left[(r_l - r_l')^2 + (z_l - z_l')^2\right].$$

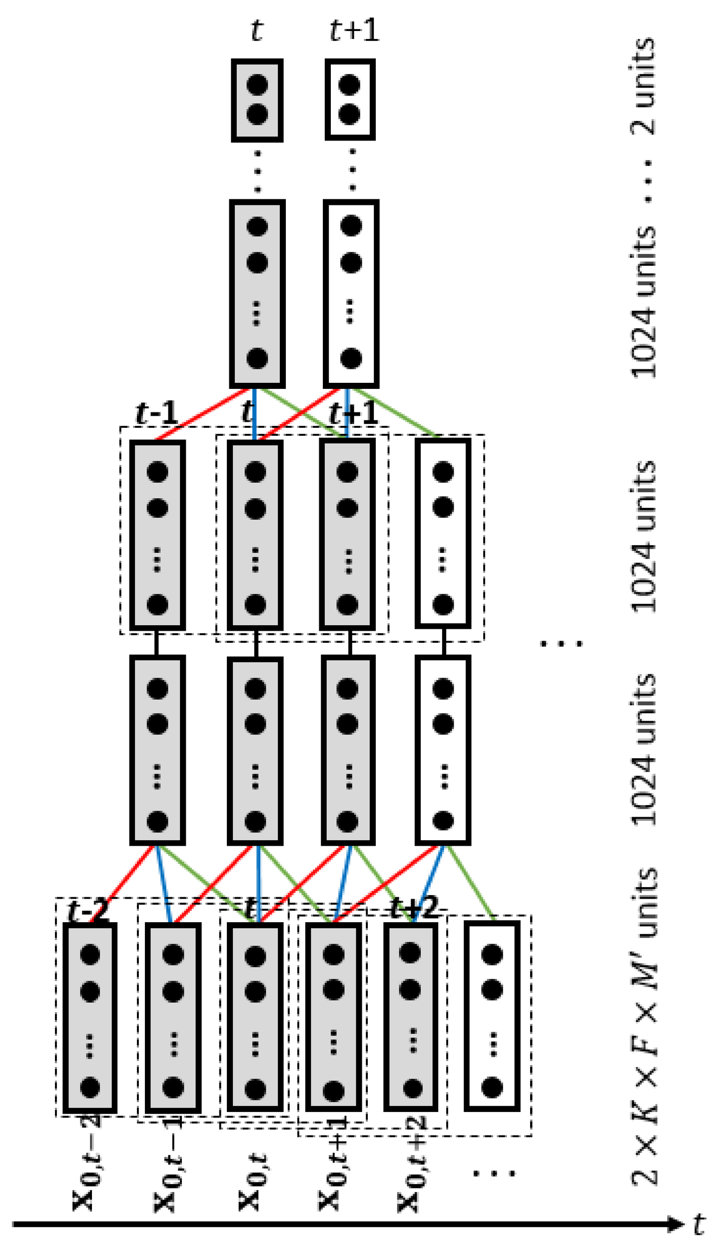

Fig. 2. Architecture of TDNN

Reference:

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 37(3), 328--339 ({1989}).

Evaluations

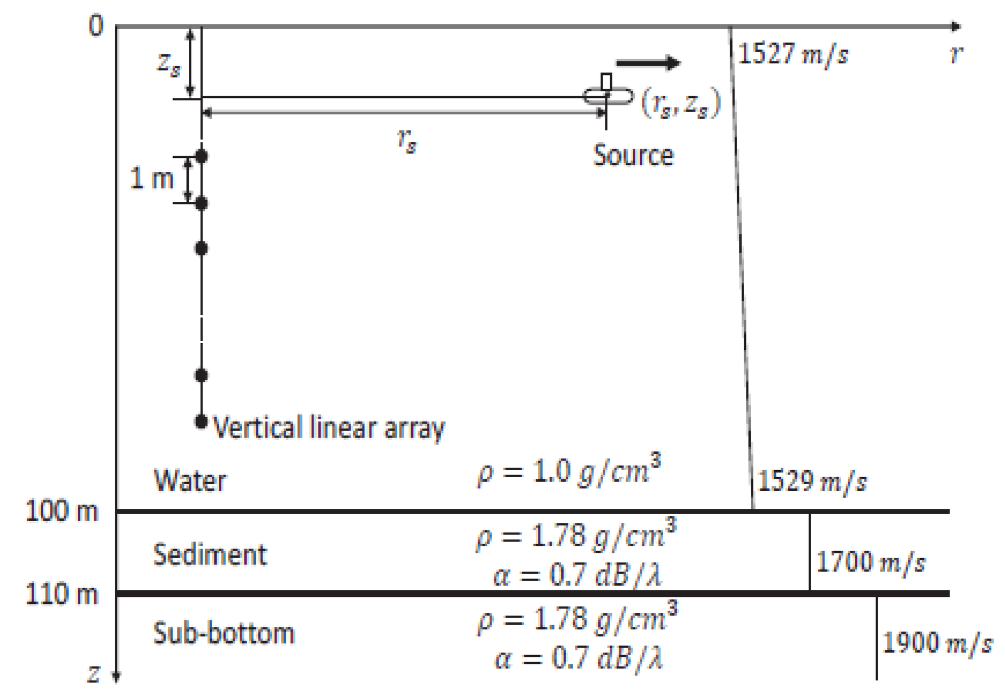

Fig. 3. Schematic diagram of the simulated acoustic environmental model

- Signal bandwidth: 50-1000Hz
- Source level: 120 dB
- Noise level: 25, 45, 65 dB
- Frame length: 0.68 s
- Feature extraction bandwidth: 100-300 Hz
- Competing method: MFP
- Toolbox: Kaldi

1) Simulation

Table 1. MAE and MRE comparison under different NLs.

Method	NL (dB)	Depth (m)	Range (km)
TDNN	25	**0.34**	**0.04 (0.2%)**
	45	**0.39**	**0.14 (0.7%)**
	65	**0.55**	**0.18 (1.0%)**
MFP	25	1.13	0.07 (0.3%)
	45	1.31	0.17 (0.7%)
	65	11.3	8.02 (43.7%)

TDNN outperforms MFP in all test condition.

2) Experiment

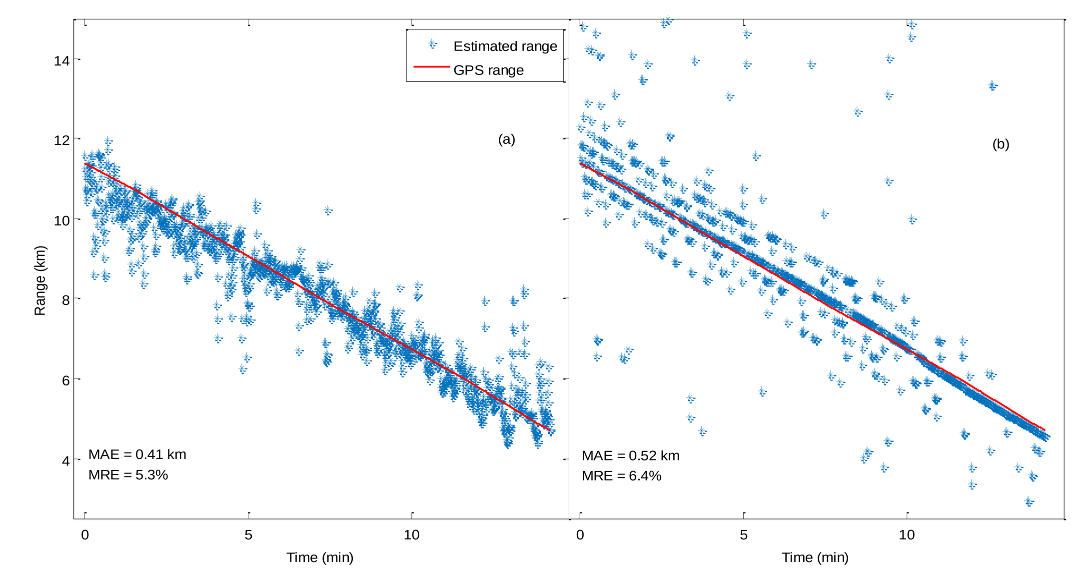

Fig. 4. Source ranging using the experimental data. (a) shows the result of the feature based method and (b) shows the result of MFP.

Table 2. MAE and MRE of range estimation for experimental data.

Method	MAE	MRE
TDNN	**0.41**	**5.3%**
MFP	0.52	6.4%

The results demonstrate that simulation data is helpful when training data are insufficient. The model trained by simulation data can also achieve a fairly good performance on experimental data.

Conclusions

In summary, our contributions are two-fold: (i) We applied TDNN to source localization task. Because DNN is a data-driven technique independent of environmental parameters, it does not rely on prior knowledge of environmental parameters and exhibits a better robustness than MFP in adverse situations for its strong nonlinear representation ability. (ii) Simulation data are available for source localization when laking of real environment training data. Simulation data in close environments can be merged to train a general model.