Cross power spectral density based beamforming for underwater acoustic communications

Jianghui Li [a], Yechao Bai [b], Youwen Zhang [c,*], Fengzhong Qu [d], Yan Wei [d], Junfeng Wang [e]

[a] *Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, UK*
[b] *The School of Electronic Science and Engineering, Nanjing University, China*
[c] *The College of Underwater Acoustic Engineering, Harbin Engineering University, Harbin, China*
[d] *Ocean College, Zhejiang University, China*
[e] *The School of Electrical and Electronic Engineering, Tianjin University of Technology, China*

ARTICLE INFO

ABSTRACT

In underwater acoustic (UWA) communications, beamforming is often used to improve the detection performance of a receiver. For beamforming, there have been methods presented in time domain, e.g. fractional delay (FD) method, and in frequency domain, e.g. time-frequency-time with cross spectral density matrix (TFT-CSDM) method. The former brings accurate direction of arrival (DOA) estimation but with high complexity and is vulnerable to noise; while the latter brings less accuracy but with lower complexity. In this paper, we propose and investigate a time-frequency-time with cross power spectral density (TFT-CPSD) beamforming method for a vertical linear array (VLA) of hydrophones. The proposed method is compared with the FD and the TFT-CSDM methods in a receiver designed for guard-free orthogonal frequency-division multiplexing (OFDM) with superimposed data and pilot signals. The comparison is based on data obtained in sea trials at distances 30–50 km in the northwest Pacific Ocean. The results demonstrate that the proposed TFT-CPSD method possesses higher accuracy than the TFT-CSDM method, and lower complexity than the FD method. Besides, the OFDM receiver with the TFT-CPSD beamforming outperforms a receiver with the TFT-CSDM beamforming and the FD beamforming at signal to noise ratio (SNR) −14 to 14 dB. The proposed beamforming technique possesses the merits of energy conservative and energy leakage reduction, which can also be applied to single-carrier transmission.

1. Introduction

In underwater acoustic (UWA) communication channels, ambient noise (e.g. radiated from sea surface wave agitation, shipping, snapping shrimps, etc.) is one of the dominant factors that affects the performance of data transmission (Wang et al., 2011; Zhou and Wang, 2014; Li et al., 2019a, 2019b, 2020; Berger et al., 2009; Zhao et al., 2016). To reduce such negative effect and improve the signal to noise ratio (SNR), receivers with vertical linear arrays (VLAs) have been developed associated with using different beamforming techniques, and are currently used in UWA communications (Xu et al., 2007; Ijaz et al., 2012; Li and Zakharov, 2018; Li et al., 2019c).

In recent decades, these beamforming techniques have been verified as of providing significant improvement in the detection performance of a receiver (Yang, 2007; Li and Zakharov, 2018). Typically, these beamforming techniques involve steps of estimating direction of arrival

(DOA), and applying such estimates to produce angle-specific directional signals for equalization and demodulation (Xu et al., 2007; Ijaz et al., 2012). The result of DOA estimation reveals the detection accuracy, and the angle-specific directional signals usually possess higher SNR than the data received directly from the acoustic channel.

There have been beamformed techniques presented in two ways, i.e. in time domain, and in frequency domain. Time domain beamforming techniques have been proved as possessing high accuracy, but they usually need interpolation between data samples, which results in high computational complexity. Moreover, such interpolation may lead to energy leakage (Hoshuyama et al., 1999), especially at low SNRs, which can reduce the detection accuracy of a DOA estimator and the detection performance of a receiver.

To reduce the complexity and remove the beamforming leakage, a frequency domain low complexity beamforming was presented in (Li and Zakharov, 2018), which is known as time-frequency-time (TFT)

---

* Corresponding author.
E-mail address: zhangyouwen@hrbeu.edu.cn (Y. Zhang).

beamforming. Such TFT beamforming firstly divides the received data packet into multiple frames, transforms the data frames from time domain to frequency domain, computes cross spectral density matrix (CSDM) for each frame in frequency domain thus providing weights to improve the beamforming performance, and finally transforms the frequency domain signal back to the time domain. Such process avoids the time domain interpolation thus reducing the computational complexity. However, such TFT beamforming applies the CSDM to compute the beamformer weights, which does not provide strong weights to the DOA as the computation of CSDM does not fully reveal the energy of data received from all directions. As a result, the accuracy of DOA estimator based on it has been presented as significantly inferior to that of the interpolation based fractional delay (FD) beamforming (Xu et al., 2007; Li and Zakharov, 2018).

The continuous time domain process of the FD beamforming does not need to separate a continuous received signal into blocks, which avoids information loss and interference from the tail with delayed signals between blocks. However, the interpolation used in such time domain process can result in high computational complexity. Besides, it may introduce another issue of beamforming energy leakage revealing at specific directions, especially at low SNR, which will be investigated in this work.

In this paper, we propose and investigate a beamforming algorithm, which provides high accuracy of DOA estimation for UWA communications utilizing a VLA of hydrophones. The proposed beamforming method computes cross power spectral density (CPSD) to estimate the beamforming weights. The investigation is based on sea trials with guard-free orthogonal frequency-division multiplexing (OFDM) signals received by a 14-element VLA (Zakharov and Kodanev, 1994; Zakharov and Morozov, 2015). The sea trials were conducted in the northwest Pacific Ocean, with a transducer towed by a vessel moving at speeds of 8 m/s and 3 m/s, at 30 km and 50 km away from the VLA, respectively. In the sea trials, the DOA estimator using the proposed TFT-CPSD beamforming shows higher accuracy than the TFT-CSDM beamforming and lower complexity than the FD beamforming. Besides, the receiver using the proposed TFT-CPSD beamforming outperforms that of using the FD beamforming and the TFT-CDSM beamforming at SNRs −14 to 14 dB. Further, we verify low beamforming energy leakage of the proposed beamforming method while showing high DOA estimation accuracy by using the Waymark propagation model simulation (Liu et al., 2012; Henson et al., 2014).

The paper is organised as follows. Section 2 describes the transmitted signal and the receiver. Sections 3 describes the spatial filter used in the receiver, and the three beamforming techniques used in the spatial filter. Section 4 compares the accuracy of DOA estimator utilizing the beamforming techniques. Section 5 compares performance of the receiver with the three beamforming techniques using sea trial data. Section 6 uses the Waymark model based simulation to verify low energy leakage from the proposed beamforming. Section 7 summarizes the paper with discussion.

2. Transmitted signal, and receiver

In this section, we consider the guard-free OFDM signals as the transmission signal, as its ability to cope with severe underwater channel conditions without complex equalization filters. The equalizer used in the receiver here is based on that presented in (Zakharov and Morozov, 2015). The receiver composes a spatial filter, in which the proposed beamforming algorithm is implemented.

2.1. Transmitted signal

The transmitted signals s(t) consists of L guard-free OFDM symbols with superimposed data and pilot signals (Zakharov and Morozov, 2015), with each OFDM symbol given by:

$$s_l(t) = \Re \left\{ e^{j2\pi f_c t} \sum_{k=-N_s/2}^{N_s/2-1} [M_1(k) + jD_l(k)] e^{j\frac{2\pi}{T_s}kt} \right\} \quad (1)$$

where $\Re\{\cdot\}$ denotes the real part of a complex number, $N_s = 1024$ is the number of sub-carriers, $f_c = 3072$ Hz is the carrier frequency, $T_s = 1$ s is the symbol duration resulting a subcarrier spacing of 1 Hz, and $j = \sqrt{-1}$. The sequence $M_1(k) \in \{-1, +1\}$ is a binary pseudo-random sequence of length N_s, serving as the pilot signal. The binary sequence $D_l(k)$ represents the information data in the lth symbol, $l = 1, 2, ..., L$, which is obtained by encoding and interleaving the original data across subcarriers using 1/2 rate convolutional code (Li et al., 2018a).

2.2. Receiver

Fig. 1(a) shows the block diagram of the receiver.

The analogue signals received by M hydrophones are bandpass filtered within the frequency bin of the OFDM transmission and converted into the digital form $r_1(i)$ to $r_M(i)$ at a sampling rate f_s, i being the discrete time index; here we apply $f_s = 4f_c = 12288$ Hz. The digital signals $r_1(i)$ to $r_M(i)$ are processed in a spatial filter that produces directional signals. In this paper, we only consider the directional signal with the highest power (see (Li and Zakharov, 2018) for maximum ratio combining technique of multiple directional signals), denoting it as $r(i, \hat{\theta})$. The DOA $\hat{\theta}$ is chosen from the average signal power as a function of received angle. The directional signal is Doppler estimated and compensated, and then equalized in time domain (Zakharov and Morozov, 2015), and transformed into the frequency domain using the fast Fourier transform (FFT). The frequency domain signal $\tilde{X}_1(k)$ is transferred to a demodulator and, after deinterleaving, further to the soft-decision Viterbi decoder (Proakis, 1995).

3. Spatial filter

Fig. 1(b) shows the block diagram of the spatial filter used in the receiver. The DOA estimator computes the spatial power distribution to estimate DOA. Then, the beamformer applies the DOA estimate to produce directional signal $r(i, \hat{\theta})$.

In the spatial filter, the following three beamforming techniques are considered:

- the FD beamforming (Xu et al., 2007; Li and Zakharov, 2018) (Section 3.1);
- the TFT-CSDM beamforming (Li and Zakharov, 2018) (Section 3.2);
- the proposed TFT-CPSD beamforming (Section 3.3).

3.1. Fractional delay (FD) beamforming

Spatial filter using FD beamforming provides accurate DOA estimation but has high complexity (Li and Zakharov, 2018). To achieve the high accuracy when processing wideband signals, such as communication signals, both the DOA estimator and beamformer should operate by introducing delays (fractional delays with respect to the sampling interval) in the hydrophone signals. The pseudo code for the FD beamforming is shown in **Algorithm 1**.

Algorithm 1 Fractional delay (FD) beamforming.

Require: hydrophone positions, received data package $r_m(nT)$ at each hydrophone m, $n = 1, 2, ..., N$;

1 **procedure**
2 **for** each interested direction θ **do**
3 compute delay $\varsigma(m, \theta) = \frac{D(m)\sin(\theta)}{c}$
4 compute $M \times N$ snapshot matrix $[\mathbf{X}(\theta)]_{m,n} = r_m(nT - \varsigma(m, \theta))$
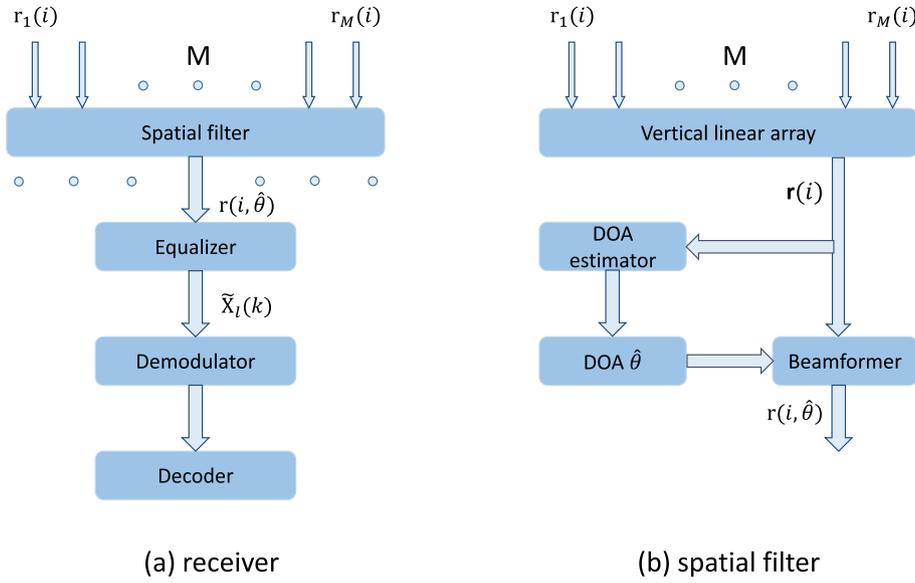
Fig. 1. (a) Block diagram of the receiver; *M* is the number of hydrophone channels (Li and Zakharov, 2018). (b) Block diagram of the spatial filter.

5 calculate sample covariance matrix $\mathbf{R}(\theta) = \mathbf{X}(\theta)\mathbf{X}^T(\theta) + \kappa\mathbf{I_M}$

6 compute spatial signal power $\tilde{\mathrm{P}}(\theta) = [\sum_{m=1}^{M}\sum_{n=1}^{M}[\mathbf{R}^{-1}(\theta)]_{m,n}]^{-1}$

7 **end for**

8 find maximum power $P_{\max} = \max_\theta \tilde{\mathrm{P}}(\theta) \to \widehat{\theta}$

9 compute weight factor $\tilde{\mathrm{W}}(\widehat{\theta}) = \tilde{\mathrm{P}}(\widehat{\theta})[\sum_{n=1}^{M}[\mathbf{R}^{-1}(\widehat{\theta})]_{1,n}, ...,$
$\sum_{n=1}^{M}[\mathbf{R}^{-1}(\widehat{\theta})]_{m,n}]^T$

10 beamformed signal $\mathrm{r}(i,\widehat{\theta}) = [\mathbf{X}^T(\widehat{\theta})\tilde{\mathrm{w}}(\widehat{\theta})]_i$

11 **end procedure**

3.1.1. FD DOA estimator

In the FD beamforming, the $M \times N$ snapshot matrix $\mathbf{X}(\theta)$ for a specific direction θ is used for calculating the diagonally loaded sample covariance matrix (Xu et al., 2007)

$$\mathbf{R}(\theta) = \mathbf{X}(\theta)\mathbf{X}^T(\theta) + \kappa\mathbf{I_M}, \tag{2}$$

and

$$[\mathbf{X}(\theta)]_{m,n} = \mathrm{r}_m(nT - \varsigma(m,\theta)), n = 1, 2, ..., N, \tag{3}$$

where $(\cdot)^T$ denotes the transpose, $\mathbf{I}_M$ is an $M \times M$ identity matrix, and κ is a loading factor which is a small positive number used here to prevent numerical instability. In our numerical examples, the value of N is set to the total number of received samples in a communication session. The signal values $\mathrm{r}_m(nT - \varsigma(m,\theta))$ in (3) are recovered by interpolation of the digital signal $\mathrm{r}_m(i)$ from the mth hydrophone at time instants $t = nT - \varsigma(m,\theta)$, where $T = 1/f_s$; for this purpose, we use the linear interpolation. The delays are different for each direction θ, computed as

$$\varsigma(m,\theta) = \frac{\mathrm{D}(m)\sin(\theta)}{c}, \tag{4}$$

where $\mathrm{D}(m)$ is the distance between the first ($m = 1$) and the mth hydrophone, and c is the sound speed . The spatial signal power $\tilde{\mathrm{P}}(\theta)$ is computed according to

$$\tilde{\mathrm{P}}(\theta) = \left[\sum_{m=1}^{M}\sum_{n=1}^{M}[\mathbf{R}^{-1}(\theta)]_{m,n}\right]^{-1} \tag{5}$$

In our numerical results, a direction grid in the interval $\theta \in [-25°, 25°]$ with a step of $0.1°$ is used.

3.1.2. FD beamformer

The beamforming weights for a direction θ are computed as

$$\tilde{\mathrm{W}}(\widehat{\theta}) = \tilde{\mathrm{P}}(\widehat{\theta})\left[\sum_{n=1}^{M}[\mathbf{R}^{-1}(\widehat{\theta})]_{1,n}, ..., \sum_{n=1}^{M}[\mathbf{R}^{-1}(\widehat{\theta})]_{m,n}\right]^T \tag{6}$$

The static DOA $\widehat{\theta}$ is estimated from the peak of the power $\tilde{\mathrm{P}}(\theta)$ for an entire communication session. The received signal for a DOA $\widehat{\theta}$ is then computed as

$$\mathrm{r}(i,\widehat{\theta}) = [\mathbf{X}^T(\widehat{\theta})\tilde{\mathrm{w}}(\widehat{\theta})]_i. \tag{7}$$

3.1.3. FD complexity

The FD beamforming technique uses interpolation and processes each direction separately, making this spatial filter complicated (Xu et al., 2007).

The DOA estimator requires the interpolation in (3), computation of the covariance matrix in (2), and the power computation in (5); complexity of the other processing can be ignored. The complexity of these three steps is given by $4N_\theta M f_s$, $4N_\theta M^2 f_s$, and $4N_\theta M^2$ multiply-accumulate operations (MACs) per second, respectively; M is the number of antenna elements, and N_θ is the number of directions in the direction grid. In the beamformer, the weight computation in (6) needs to be performed; the other operations require significantly lower complexity. This step requires $(4M^2 f_s + 2M f_s)$ MACs. For example, with $M = 14$, $N_\theta = 501$, and $f_s = 12288$ Hz, i.e. with parameters used in the receiver in Section 2.2, the total complexity of the spatial filter is 5.2×10^9 MACs.

3.2. TFT-CSDM beamforming

Spatial filter using the TFT-CSDM beamforming first divides the continuous time domain received signals into L frames, and then transfers these time domain frames into frequency domain for DOA estimation and beamforming. Finally, the weighted frequency domain signal is transformed back into time domain. The pseudo code for the TFT-CSDM beamforming is shown in **Algorithm 2**.

Algorithm 2 TFT-CSDM beamforming.

Require: hydrophone positions, received signals $\mathbf{r}(i) = [\mathrm{r}_1(i), ..., \mathrm{r}_M(i)]^T$, frequency bin width $\Delta\omega$;

Ensure: frequency bandwidth F, frequency bin number $K = 2\pi F/\Delta\omega$, interested direction θ;

1 **procedure**
2 **for** $k = 0, 1, ..., K - 1$ **do**
3 compute frequency domain snapshot
   $\mathbf{z}(i_l; k) = \sum_{n=0}^{I_L-1} \mathbf{r}(i_l + n)e^{-j\omega_k n/f_s}$
4 compute CSDM $\mathbf{Y}(i_l; k) = \frac{1}{L}\sum_{l=1}^{L}\mathbf{z}(i_l + (l - 1)I_L; k)\mathbf{z}^*(i_l + (l - 1)I_L;$
   $k) + \kappa\mathbf{I}_M$
5 compute steering vector

   $\mathbf{v}(\theta, k) = \left[1, ..., e^{-j\omega_k\frac{D(m)\sin(\theta)}{c}}, ..., e^{-j\omega_k\frac{D(M)\sin(\theta)}{c}}\right]$

6 compute power $\mathrm{P}_k(i_l; \theta) = [\mathbf{v}^H(\theta, k)\mathbf{Y}^{-1}(i_l; k)\mathbf{v}(\theta, k)]^{-1}$
7 **end for**
8 compute average power $\tilde{\mathrm{P}}(\theta) = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=0}^{K-1}\mathrm{P}_k(i_l; \theta)$
9 find maximum power $P_{\max} = \max_m P_m \rightarrow \widehat{\theta}$
10 **for** $k = 0, 1, ..., K - 1$ **do**
11 compute weight factor $\overline{\mathbf{w}}_l(\widehat{\theta}, k) = \mathbf{Y}^{-1}(i_l; k)\mathbf{v}(\widehat{\theta}, k)\mathrm{P}_k(i_l; \widehat{\theta})$
12 smooth weight factor $\mathbf{w}_l(\widehat{\theta}, k) \leftarrow \lambda\mathbf{w}_{l-1}(\widehat{\theta}, k) + (1 - \lambda)\overline{\mathbf{w}}_l(\widehat{\theta}, k)$
13 beamformed signal $\mathrm{r}(i, \widehat{\theta}) = \sum_{k=0}^{K-1}\mathbf{w}_l^*(\widehat{\theta}, k)\mathbf{z}(i_l; k)e^{j\omega_k n/f_s}$
14 **end for**
15 **end procedure**

3.2.1. TFT-CSDM DOA estimator

The DOA estimator computes the spatial power distribution of the received signal by processing the hydrophone signals $\mathrm{r}_1(i)$ to $\mathrm{r}_M(i)$. The ith time domain snapshot of the received signals is described as an $M \times 1$ vector $\mathbf{r}(i) = [\mathrm{r}_1(i), ..., \mathrm{r}_M(i)]^T$. The snapshots are divided into L frames of I_L snapshots each. The $M \times 1$ frequency domain snapshot at frequency ω_k for a frame starting at time instant i_l is given by

$$\mathbf{z}(i_l; k) = \sum_{n=0}^{I_L-1} \mathbf{r}(i_l + n)e^{-j\omega_k n/f_s}, \tag{8}$$

where $k = 0, 1, ..., K - 1$, $K = 2\pi F/\Delta\omega$, F is the bandwidth, $\omega_k = \omega_0 + k\Delta\omega$, $\Delta\omega = 2\pi\Delta f$, and ω_0 is the lowest frequency ($\omega_0 = 2\pi(f_c - F/2)$ here). For a frame starting at time instant i_l, for every frequency ω_k, the $M \times M$ cross spectral density matrix (CSDM) is computed as (Haykin et al., 1985):

$$\mathbf{Y}(i_l; k) = \frac{1}{L}\sum_{l=1}^{L}\mathbf{z}(i_l + (l-1)I_L; k)\mathbf{z}^*(i_l + (l-1)I_L; k) + \kappa\mathbf{I}_M, \tag{9}$$

where $(\cdot)^*$ denotes the conjugate transpose, l is the frame index, and κ is a loading factor which is a small positive number related to the noise level. In the experiments, the loading factor κ was set to a small value to prevent numerical instability when inverting the matrix $\mathbf{Y}(i_l; k)$ (see below). More specifically, it was set to at most 10^{-8} of $(1/M)\mathrm{trace}\{\mathbf{Y}(i_f; k)\}$, where $\mathrm{trace}\{\cdot\}$ is the matrix trace. The loading factor κ can be optimized to achieve an improved detection performance (Li et al., 2003), while such optimization is not detailed here. The matrix $\mathbf{Y}(i_l; k)$ is used for obtaining the spatial power at each direction θ.

For beamforming, various algorithms have been presented in literature, e.g. conventional classic beamforming (Sutton, 1979; Xu et al., 2020; Ozanich et al., 2020), minimum norm beamforming (MINNORM), Multiple Signal Classification Algorithm (MUSIC), root-MUSIC, Estimation of signal parameters via rotation invariance techniques (ESPRIT), minimum variance distortionless response algorithm (MVDR), etc. (Raghunath and Reddy, 1992; Vaidyanathan and Buckley, 1995; Swingler, 1999). The classic beamforming algorithm does not provide high resolution (Quinquis and Boulinguez, 1997), while the MINNORM, MUSIC, root-MUSIC, and ESPRIT algorithms (Bai et al., 2019) are able to provide high resolution on the DOA estimation. However, these high resolution algorithms are often based on the computation of inverse QR-based decomposition, which requires extra

computing resources. The matrix inversion unit of the decomposition only works for a fixed set of matrix (Ma et al., 2011), which limits the implementation of these high resolution algorithms in UWA communications with long data sets. Thus we choose an algorithm without using the QR-based decomposition, i.e. MVDR algorithm (Capon, 1969; Alexander, 1986) to compute the spatial power.

For a frequency ω_k, the steering vector is given by

$$\mathbf{v}(\theta, k) = \left[1, ..., e^{-j\omega_k\frac{D(m)\sin(\theta)}{c}}, ..., e^{-j\omega_k\frac{D(M)\sin(\theta)}{c}}\right]. \tag{10}$$

The power at frequency ω_k from a direction θ is given by:

$$\mathrm{P}_k(i_l; \theta) = \left[\mathbf{v}^H(\theta, k)\mathbf{Y}^{-1}(i_l; k)\mathbf{v}(\theta, k)\right]^{-1}, \tag{11}$$

and the total power for all frequencies of interest is given by

$$\mathrm{P}(i_l; \theta) = \sum_{k=0}^{K-1}\mathrm{P}_k(i_l; \theta). \tag{12}$$

The average power over L frames is given by

$$\tilde{\mathrm{P}}(\theta) = \frac{1}{L}\sum_{l=1}^{L}\mathrm{P}(i_l; \theta). \tag{13}$$

3.2.2. TFT-CSDM beamformer

In this case, the DOA $\widehat{\theta}$ is chosen from the peak of the $\tilde{\mathrm{P}}(\theta)$ for an entire communication session. For a chosen DOA $\widehat{\theta}$, to cancel the interference arriving from other directions, the beamformer weight vector $\overline{\mathbf{w}}_l(\widehat{\theta}, k)$ of the lth frame is calculated as (Capon, 1969):

$$\overline{\mathbf{w}}_l(\widehat{\theta}, k) = \mathbf{Y}^{-1}(i_l; k)\mathbf{v}(\widehat{\theta}, k)\mathrm{P}_k(i_l; \widehat{\theta}). \tag{14}$$

The weight vector is then smoothed in time:

$$\mathbf{w}_l(\widehat{\theta}, k) \leftarrow \lambda\mathbf{w}_{l-1}(\widehat{\theta}, k) + (1 - \lambda)\overline{\mathbf{w}}_l(\widehat{\theta}, k), \tag{15}$$

where $0 \leq \lambda < 1$ is a forgetting factor, and $\mathbf{w}_0(\widehat{\theta}, k) = \overline{\mathbf{w}}_1(\widehat{\theta}, k)$. The directional signal is then computed as:

$$\mathrm{r}(i, \widehat{\theta}) = \sum_{k=0}^{K-1}\mathbf{w}_l^*(\widehat{\theta}, k)\mathbf{z}(i_l; k)e^{j\omega_k n/f_s}, \tag{16}$$

where $i = i_l + (l - 1)I_L + n$.

3.2.3. TFT-CSDM complexity

For the DOA estimation, the spatial filter requires the time-frequency transform in (8), computation of the CSDM in (9), and the power computation in (11); complexity of the other processing can be ignored. The complexity of these three steps is given by $2KMf_s$, $4Kf_sM^2/I_L$, and $4(KM^3 + KN_\theta M^2)$ MACs per second, respectively; N_θ is the number of directions in the direction grid. In a beamformer, the frequency-time transform in (16) needs to be performed; the other operations require significantly lower complexity. This step requires $(4KMf_s/I_L + 4Kf_s)$ MACs. For example, with $M = 14$, $K = 16$, $f_s/I_L = 1$, $N_\theta = 501$, and $f_s = 12288$ Hz, i.e. with parameters used in the receiver in Section 2.2, the total complexity of the spatial filter is 1.3×10^7 MACs.

3.3. Proposed TFT-CPSD beamforming

When the gradient of sound is significant along the array aperture, the wave front is not spherical and beamforming should be replaced with a mode filtering, otherwise energy leakage cannot be avoided. The continuing processing of signal block (frame) with different delays cannot be done by blocks and FFT, because of the tail with delayed signals. Thus, convolution methods, overlap-save or overlap-add methods are needed (Muramatsu and Kiya, 1997; Daher et al., 2010).

Different from the existing TFT-CSDM beamforming, the TFT-CPSD beamforming first applies the overlap-save method for a frame length I_L, then computes the cross power spectral density (CPSD) instead of CSDM for each segment to obtain spatial power, uses window (Hamming window here) to filter data for each frame, and smooths the weight vector using a moving average filter on neighbour frames instead of that from the former frames initiated from the first frame as shown in (15). The pseudo code for the TFT-CPSD beamforming is concluded in **Algorithm 3**.

Algorithm 3 TFT-CPSD beamforming.

Require: hydrophone positions, received signals $\mathbf{r}(i) = [\mathbf{r}_1(i), \ldots, \mathbf{r}_M(i)]^T$, frequency bin width $\Delta\omega$;

Ensure: frequency bandwidth F, frequency bin number $K = 2\pi F/\Delta\omega$, interested direction θ;

  1 **procedure**
  2 reconstruct each data frame $\mathbf{r}(i_l)$ with the overlap-save method
  3 **for** $k = 0, 1, \ldots, K-1$ **do**
  4 compute cross-correlation sequence $\tilde{\mathbf{R}}(i_l; m) = E\{\mathbf{r}(i_l + n + m)\mathbf{r}^*(i_l + n)\}$
  5 compute CPSD $\tilde{\mathbf{C}}(i_l, k) = \sum_{m=-I_L}^{I_L} \tilde{\mathbf{R}}(i_l; m)e^{-j\omega_k m}$
  6 compute steering vector $\mathbf{v}(\theta, k) = \left[1, \ldots, e^{-j\omega_k \frac{D(m)\sin(\theta)}{c}}, \ldots, e^{-j\omega_k \frac{D(M)\sin(\theta)}{c}}\right]$
  7 compute power $P_k(i_l; \theta) = [\mathbf{v}^*(\theta, k)\tilde{\mathbf{C}}(i_l, k)\mathbf{v}(\theta, k)]^{-1}$
  8 **end for**
  9 compute average power $\tilde{P}(\theta) = \frac{1}{L}\sum_{l=1}^{L}\sum_{k=0}^{K-1} P_k(i_l; \theta)$
  10 find maximum power $P_{\max} = \max_m P_m \rightarrow \hat{\theta}$
  11 **for** $k = 0, 1, \ldots, K-1$ **do**
  12 compute weight factor $\overline{\mathbf{w}}_l(\hat{\theta}, k) = \tilde{\mathbf{C}}^{-1}(i_l; k)\mathbf{v}(\hat{\theta}, k)P_k(i_l; \hat{\theta})$
  13 smooth weight factor $\mathbf{w}_l(\hat{\theta}, k) \leftarrow \frac{\overline{\mathbf{w}}_{l-l_d}(\hat{\theta}, k) + \ldots + \overline{\mathbf{w}}_l(\hat{\theta}, k) + \ldots + \overline{\mathbf{w}}_{l+l_d}(\hat{\theta}, k)}{2l_d + 1}$
  14 compute frequency domain snapshot $\mathbf{z}(i_l; k) = \sum_{n=0}^{I_L - 1} \mathbf{r}(i_l + n)e^{-j\omega_k n/f_s}$
  15 beamformed signal $\mathbf{r}(i, \hat{\theta}) = \sum_{k=0}^{K-1} \mathbf{w}_l^*(\hat{\theta}, k)\mathbf{z}(i_l; k)e^{j\omega_k n/f_s}$
  16 **end for**
  17 **end procedure**

3.3.1. TFT-CPSD DOA estimator

The DOA estimator computes the spatial power distribution of the received signal by processing the hydrophone signals $\mathbf{r}_1(i)$ to $\mathbf{r}_M(i)$. The ith time domain snapshot of the received signals is described as an $M \times 1$ vector $\mathbf{r}(i) = [\mathbf{r}_1(i), \ldots, \mathbf{r}_M(i)]^T$.

The snapshots are then divided into L frames of I_L snapshots each. Different from the TFT-CSDM beamforming, here the frame is overlapped with its previous frame. Each frame has a length of I_L and has an overlap length I_{Lo} with its previous frame. Here we set the overlap length I_{Lo} as half of I_L, and will investigate the length for each frame with the experimental data presented in Section 5.1.1. For each frame, we use a Hamming window of length H_{win} to filter the data segments of that window length. For a frame starting at time instant i_l, for every frequency of interest ω_k, the CPSD is the distribution of power per unit frequency (Welch, 1967; Oppenheim et al., 2001):

$$\tilde{\mathbf{C}}(i_l, k) = \sum_{m=-I_L}^{I_L} \tilde{\mathbf{R}}(i_l; m)e^{-j\omega_k m}. \tag{17}$$

The frequency of interest ω_k is chosen from a bin vector with a bin width of $\Delta F = F/K$. For each bin, we integrate the wideband across the frequency bin width ΔF assuming that the variation in a bin can be

omitted.

The cross-correlation sequence $\tilde{\mathbf{R}}(i_l; m)$ is defined as

$$\tilde{\mathbf{R}}(i_l; m) = E\{\mathbf{r}(i_l + n + m)\mathbf{r}^*(i_l + n)\}, \tag{18}$$

where $\mathbf{r}(i_l + n)$ is the snapshots in the lth frame, $-I_L < n < I_L$ and $-I_L < m < I_L$ for a single frame, and $E\{\cdot\}$ is the expected value operator. In practice, it can be achieved by computing

$$\tilde{\mathbf{R}}(i_l; m) = \begin{cases} \sum_{n=0}^{N-m-1} \mathbf{r}(i_l + n + m)\mathbf{r}^*(i_l + n), & (m \geq 0) \\ \tilde{\mathbf{R}}^*(i_l; -m), & (m < 0) \end{cases} \tag{19}$$

with normalization to produce an accurate estimate.

The CPSD $\tilde{\mathbf{C}}(i_l, k)$ is used for obtaining the spatial power at every direction θ using the MVDR algorithm (Capon, 1969; Alexander, 1986). For a frequency ω_k, the steering vector is given by (10). The power at frequency ω_k from a direction θ is given by:

$$P_k(i_l; \theta) = [\mathbf{v}^*(\theta, k)\tilde{\mathbf{C}}(i_l, k)\mathbf{v}(\theta, k)]^{-1} \tag{20}$$

and the total power for all frequencies of interest is given by

$$P(i_l; \theta) = \sum_{k=0}^{K-1} P_k(i_l; \theta). \tag{21}$$

The average power over L frames is given by

$$\tilde{P}(\theta) = \frac{1}{L}\sum_{l=1}^{L} P(i_l; \theta). \tag{22}$$

3.3.2. TFT-CPSD beamformer

In this case, the DOA $\hat{\theta}$ is also chosen from the peak of the $\tilde{P}(\theta)$ for the entire session. For a chosen DOA $\hat{\theta}$, to cancel the interference arriving from other directions, the beamformer weight vector $\overline{\mathbf{w}}_l(\hat{\theta}, k)$ in the lth frame is calculated as (Capon, 1969):

$$\overline{\mathbf{w}}_l(\hat{\theta}, k) = \tilde{\mathbf{C}}^{-1}(i_l; k)\mathbf{v}(\hat{\theta}, k)P_k(i_l; \hat{\theta}). \tag{23}$$

Due to ocean dynamics resulting in fluctuations of DOA during the communication session, the DOA associated weights may change significantly from the beginning. Instead of using iterative smooth from the past frames as shown in (15), we introduce an average smooth, in which the weight vector is smoothed using a moving average filter. The filter uses a number of data points $l_d = \hat{\lambda}L/2$ for calculating the smoothed value. The parameter $\hat{\lambda}/2$ is in the range (0,1) denoting a fraction of the total number of data points. The weight vector is then smoothed as:

$$\mathbf{w}_l(\hat{\theta}, k) \leftarrow \frac{\overline{\mathbf{w}}_{l-l_d}(\hat{\theta}, k) + \ldots + \overline{\mathbf{w}}_l(\hat{\theta}, k) + \ldots + \overline{\mathbf{w}}_{l+l_d}(\hat{\theta}, k)}{2l_d + 1}, \tag{24}$$

where $\mathbf{w}_1(\hat{\theta}, k) = \overline{\mathbf{w}}_1(\hat{\theta}, k)$, and $\mathbf{w}_2(\hat{\theta}, k) = \frac{\overline{\mathbf{w}}_1(\hat{\theta}, k) + \overline{\mathbf{w}}_2(\hat{\theta}, k)}{2}$, etc. The directional signal is then computed as:

$$\mathbf{r}(i, \hat{\theta}) = \sum_{k=0}^{K-1} \mathbf{w}_l^*(\hat{\theta}, k)\mathbf{z}(i_l; k)e^{j\omega_k n/f_s}, \tag{25}$$

where $i = i_l + (l - 1)I_L + n$. While adding these directional signal snapshots together, we overlap the extra data length Lo to reduce the tail effect and energy leakage.

3.3.3. TFT-CPSD complexity

For the DOA estimation, the spatial filter requires the cross-correlation in (18), the computation of CPSD in (17), and the power computation in (20); complexity of the other processing can be ignored. The complexity of the cross-correlation is computed from the integration

of the number of non-zeros multiplications. The complexity of these three steps is given by $4M^2f_s(H_{win}+1)/2$, $2KMf_s$, and $4(KM^3+KN_\theta M^2)$ MACs per second, respectively. In a beamformer, the frequency-time transform in (25) needs to be performed; the other operations require significantly lower complexity. This step requires $(4KMf_s/ I_L + 4Kf_s)$ MACs. For example, with $M=14$, $K=16$, $H_{win}=16$, $N_\theta = 501$, and $f_s = 12288$ Hz, i.e. with parameters used in the receiver in Section 2.2, the total complexity of the spatial filter is 9.5×10^7 MACs.

4. Accuracy of DOA estimation

To compare the accuracy of DOA estimator and detection capability of using the three beamforming techniques, we use the data recorded in the sea trial session F1-1 at a transmitter to VLA receiver distance of 30 km (detailed in Section 5).

4.1. Parameter justification of TFT-CPSD

We first justify the values of window length and frequency bin of the proposed TFT-CPSD algorithm. Fig. 2 shows comparison results of average spatial signal power $\tilde{P}(\theta)$ of the DOA estimator at various window lengths and frequency bins. Fig. 2(a) shows that when the frequency bin $\Delta F = 16$ Hz, the DOA estimator shows the best accuracy as the window length $H_{win} = 16$ samples; and Fig. 2(b) shows that when the window length $H_{win} = 16$ samples, the accuracy of DOA estimator increases as the frequency bin ΔF decreases, while it is almost unchanged as $\Delta F \leq 64$ Hz.

Thus we choose $H_{win} = 16$ and $\Delta F = 64$ Hz for further processing. When processing the received signals in the spatial filter, $K = 1024/ 64 = 16$ frequencies are processed in the bandwidth of interest $F = 1024$ Hz, and the lowest frequency of interest $f_0 = \omega_0/(2\pi) = 2560$ Hz. The frame length I_f is considered to be one OFDM symbol here, and the loading factor $\kappa = 10^{-3}$. The DOAs θ for DOA estimation are computed in $[-25°, 25°]$ with a DOA step of $0.1°$.

4.2. Comparison of DOA estimators

Fig. 3 shows comparison results of average spatial signal power $\tilde{P}(\theta)$ estimated from the DOA estimator using the three beamforming techniques at different SNRs, i.e. $[-15, -5, 5, 14]$ dB, by adding measured ambient noise to received signal for each hydrophone channel. The proposed DOA estimator using the proposed TFT-CPSD beamforming outperforms that using the TFT-CSDM beamforming in accuracy, while it is inferior to that of using the FD beamforming, obvious at high SNRs. However, at low SNRs, the FD beamforming shows significant beamforming leakage of the target signal into the interference at multiple angles, which makes the accuracy of it worse and makes the DOA detection difficult.

The computational complexity of the DOA estimator using the three beamforming techniques are compared in Table 1. The complexity of the

DOA estimator using the proposed TFT-CPSD beamforming is significantly lower than that of using the FD beamforming, while it is not much higher than that of using the TFT-CSDM beamforming.

5. Receiver performance

In this section, we compare the receiver with DOA estimator using different beamforming techniques. To demonstrate the effectiveness of the proposed TFT-CPSD beamforming used in the receiver, comparisons are performed with the transmission of guard-free OFDM signals with superimposed data and pilot (Zakharov and Morozov, 2015). These comparisons use data from two sea trial sessions in the northwest Pacific Ocean, i.e.

- session F1-1: transmitter to receiver distance of 30 km;
- session F-3: transmitter to receiver distance of 50 km.

In both sessions, the depth of the transmitter was 250 m, and the depth of the first VLA hydrophone was 420 m (Fig. 4).

In the receive VLA, the distances from the mth hydrophones in turn to the first (top) hydrophone are [0 0.6 1.2 1.8 2.4 3.0 3.6 3.9 4.8 5.4 6.0 6.6 7.8 8.1] m. The sound-speed profile (SSP) measured in the sea trial area is shown in Fig. 5, showing the gradient of sound. The sea depth is about 5 km, and the minimum sound speed is at a depth of about 300 m. In the two sea trial sessions, communication signals are transmitted in the frequency band 2560–3584 Hz.

5.1. Session F1 − 1

In the F1 − 1 session, the transmitter was towed by a vessel moving towards the receiver at a high speed of 8 m/s, and the distance between them varied from 30 to 29 km. In this session, 100 guard-free OFDM symbols were transmitted. Fig. 6(a) left side shows the spatial power distribution. It can be seen that an outstanding cluster is identified as the one with DOA around $\hat{\theta} = -1.4°$. Fig. 6(a) right side shows the time-varying DOA detected for each frame (blue solid line) and a static DOA $\hat{\theta} = -1.4°$ chosen from the average spatial power peak for the entire communication session (red dashed line). The time-varying DOA across the static DOA possesses a maximum variance of $1.5°$ from ocean dynamics. The static angle $\hat{\theta} = -1.4°$ is used to produce a directional signal using the three beamforming techniques.

Fig. 7(a) shows the time-varying SNR at the first receive VLA hydrophone in the F1-1 session, which is the result of received signal energy divided by recorded noise energy in frames. The SNR varies between 7 dB and 18.5 dB, and on average is 14 dB, indicating complex noise levels in the communication channel. Fig. 8(a) shows fluctuations of the channel impulse response over the F1-1 session at the first hydrophone, revealing a single outstanding propagation path of the transmitted signal in the channel.
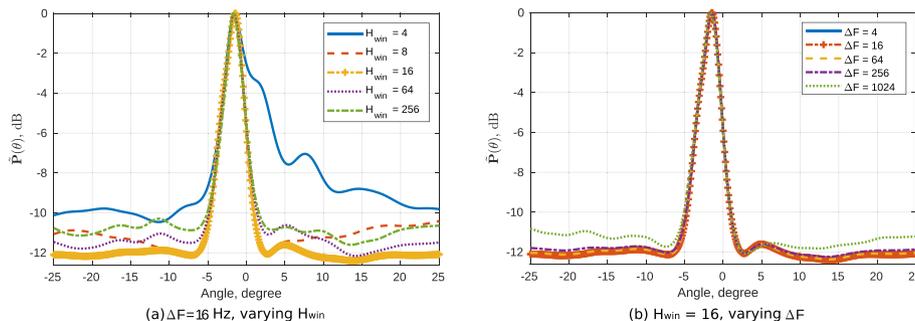
Fig. 2. Comparisons of average spatial signal power $\tilde{P}(\theta)$ estimated from the DOA estimator using the proposed TFT-CPSD beamforming with different window length H$_{win}$ and frequency bins ΔF in the sea trial session at a distance of 30 km. (a) varying window length; (b) varying frequency bin.

(a) SNR = 14 dB

(b) SNR = 5 dB

(c) SNR = -5 dB

(d) SNR = -15 dB
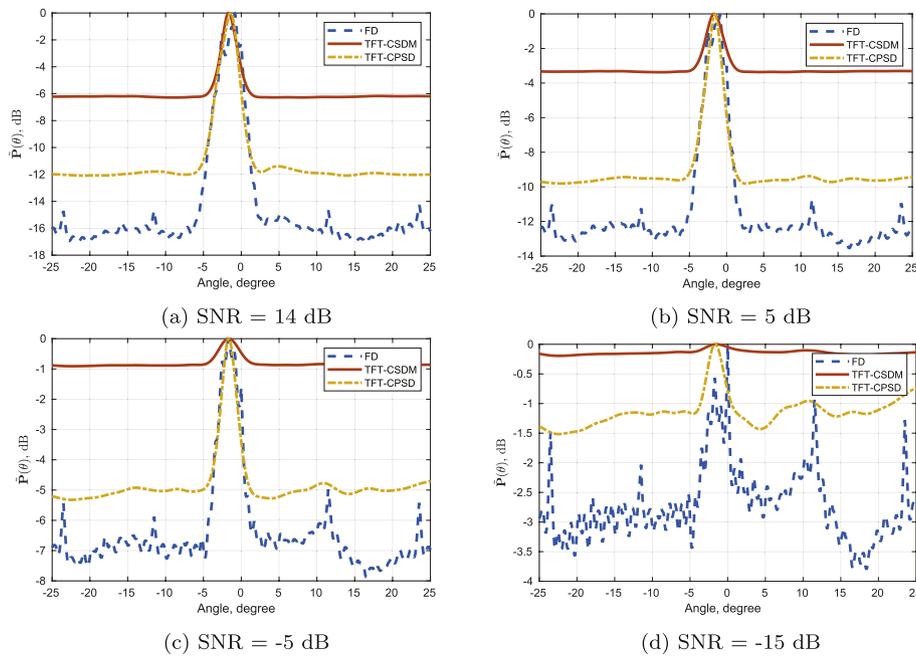
Fig. 3. Comparisons of average spatial signal power $\tilde{P}(\theta)$ estimated from the DOA estimator using the FD beamforming, the TFT-CSDM beamforming, and the proposed TFT-CPSD beamforming in the sea trial session F1-1 at different SNRs. (a) 14 dB. (b) 5 dB. (c) −5 dB. (d) −15 dB.

Table 1
Complexity of DOA estimator with different beamforming.

Beamforming	Complexity (10^6 MAC/s)
no beamforming	0
FD	5200
TFT-CSDM	13
TFT-CPSD (proposed)	95
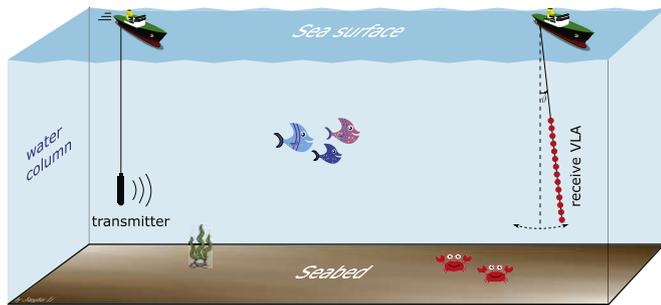
Fig. 4. Sea trial scenario in the northwest Pacific Ocean. The depth of the transmitter was 250 m, and the depth of the first receive VLA hydrophone was 420 m. The length of the receive VLA of hydrophones is 8.1 m. The receive VLA oscillation can be resulted from the ocean dynamics. Underwater ambient noise in the communication channel can be radiated from surface wave agitation, shipping, marine animals, turbulence, etc. (Pizzuti et al., 2012; Wittekind and Schuster, 2016; Hodges, 2011; Asolkar et al., 2017; Liu et al., 2005; Li et al., 2018b; Brooker and Humphrey, 2016; Kellett et al., 2013).

5.1.1. Frame length investigation

Underwater acoustic channel is often characterized as temporally and spectrally fast-varying. The variation of DOA and Doppler can be significant from one frame to the other. The continuous processing of signal frames with different delays introduces energy leakage inevitably because of the tail of delayed signals. To reduce such leakage, we investigate the optimal frame length adapting to the specific channel for the process of continuous signal in the receiver. Here we investigate the
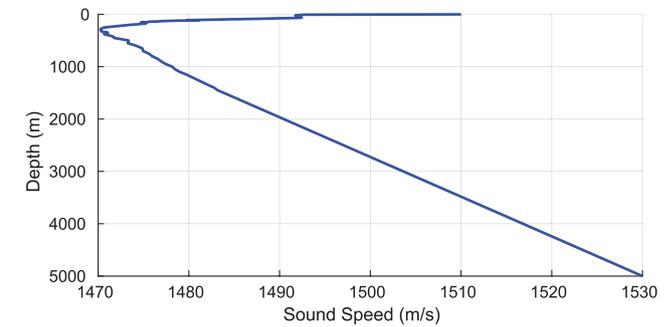
Fig. 5. Sound speed profiles (SSP) measured in the sea trial area (northwest Pacific Ocean), and used in the simulation (Section 6).

frame length in the receiver based on the data collected from the session F1-1.

In the investigation, the TFT-CPSD beamforming is implemented in the receiver, and the frame length is set to 1 s (s), 1/3 s, 1/6 s, 1/12 s, and 1/24 s, respectively. Fig. 9 shows the bit error rate (BER) performance of the receiver as different length frames are processed. It shows that when the frame length is set to 1/6 s, the receiver shows the best performance at SNR higher than 0 dB. It also shows that the receiver is sensitive to the frame length at high SNR, while it is insensitive at low SNR.

5.1.2. BER performance comparison

The BER performances of the receiver using the three beamforming techniques based DOA estimator are now compared. To show the performance of the receiver at different SNR, we add noise to the received signals separately. Signals with lower SNR are produced by adding measured ambient noise from each hydrophone to the received signal with SNR of 14 dB (Fig. 7(a)). Note that the ambient noise varies in bathymetry and the depth/position of hydrophones, resulting in specific relationship/correlations among these channel noises recorded by the 14 hydrophones (see details in Appendix A). Fig. 10 presents the BER performance of the receiver applied to the sea trial data recorded in the F1-1 session at spectral efficiencies of (a) 1 bps/Hz and (b) 0.5 bps/Hz;
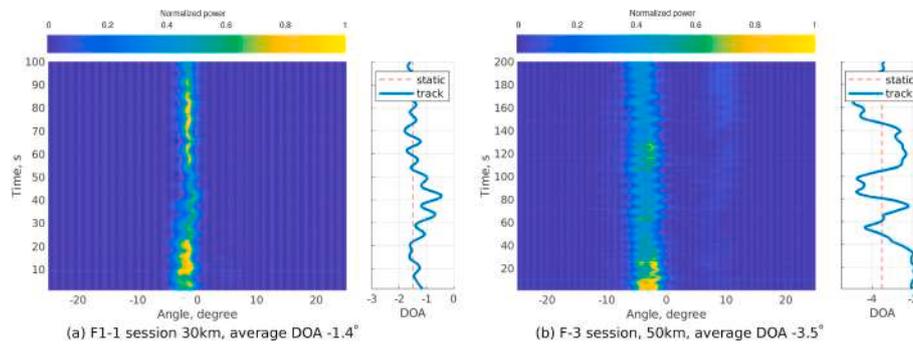
Fig. 6. DOA fluctuation in the sea trial. (a) F1-1 session, 30 km; (b) F-3 session, 50 km. Left: estimated spatial signal power; right: angle of spatial power peak for the entire session (red dashed line), and angle of spatial power peak for each data frame (blue solid line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
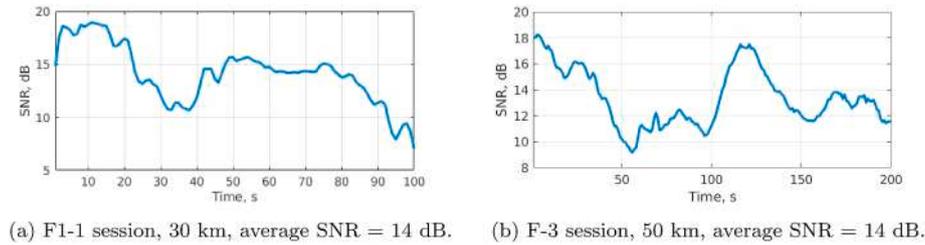
Fig. 7. Time-varying SNR at the first (top) hydrophone channel in the sea trial. (a) F1-1 session, 30 km; (b) F-3 session, 50 km.
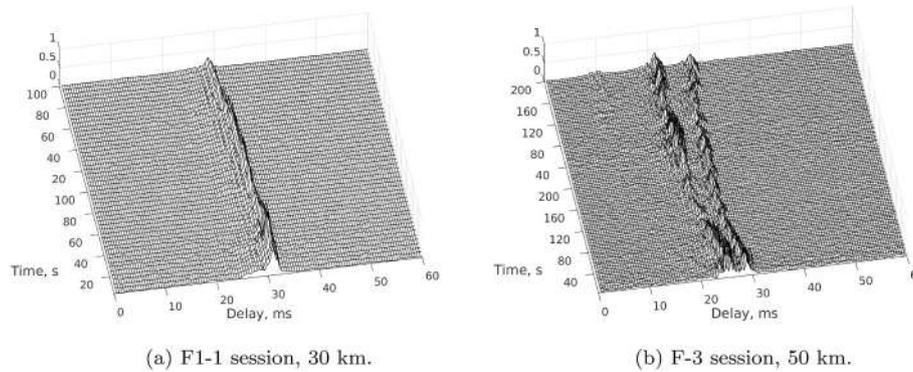
Fig. 8. Fluctuations of the channel impulse response at the first hydrophone in the two communication sessions. (a) F1-1 session; (b) F-3 session.

the convolutional code represented by polynomial in octal, being rate-1/2 code (Proakis, 1995) is used.

Results presented in Fig. 10 demonstrate that when the SNR increases from −14 dB to 14 dB, the receiver with all the three beamforming techniques show improved detection performance at both spectral efficiencies compared to that without using beamforming. The receiver using the proposed TFT-CPSD beamforming provides better performance through the entire range of SNR than the FD beamforming and the TFT-CSDM beamforming, with only slightly comparable to the FD beamforming at high SNR (> 9 dB) at spectral efficiency of 1 bps/Hz. At a lower spectral efficiency (1/2 bps/Hz), the receiver using the TFT-CPSD beamforming technique outperforms both the FD beamforming and TFT-CSDM beamforming, and achieves error-free transmission at SNR ≥ -2 dB, showing better detection performance than that of using the other two beamforming techniques. The TFT-CPSD performs better than the FD beamforming because of its reduced energy leakage between overlapped segments. The TFT-CPSD performs better than the TFT-CSDM beamforming because of its fully computed energy of the received signals.

5.1.3. Session F − 3

In the F − 3 session, the transmitter was towed by a vessel moving away from the receive VLA at a speed of 3 m/s, and the distance between them varied from 50 to 51 km. In this session, 200 guard-free OFDM symbols were transmitted. Fig. 6(b) left side shows the spatial power distribution. It can be seen that a mixed cluster, i.e. mixed by two separated sub-clusters from two time-varying arrival DOAs as observed, is identified as with DOA around $\hat{\theta} = -3.5°$. Due to the difficulty of separating the two sub-clusters as in such close angle case, we consider it as a single cluster to find the static DOA with the peak of average spatial signal power through the session (see (Li and Zakharov, 2018) for technique of processing multiple DOA branches). Fig. 6(b) right side shows the time-varying DOA detected for each frame (blue solid line) and a static DOA for the entire communication session (red dashed line). The peak of time-varying DOA changes between the two sub-clusters through the session indicates comparable strength of the two path arrivals. The time-varying DOA across the static DOA possesses a maximum variation of 3.0°. The static DOA is used to produce a single directional signal.

Fig. 7(b) shows the time-varying SNR at the first receive VLA
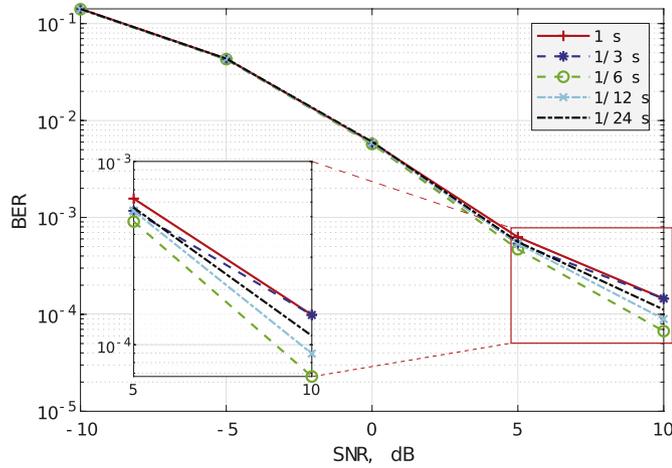
Fig. 9. BER performance of the receiver with different signal frame length (1 s, 1/3 s, 1/6 s, 1/12 s, and 1/24 s). It shows the best performance as the frame length set to 1/6 s (s).

hydrophone in the F-3 session, varying between 9 dB and 18 dB, and on average is 14 dB. Fig. 8(b) shows fluctuations of the channel impulse response over the F-3 session at the first hydrophone, revealing two outstanding path arrivals from the channel. Rather than possessing an outstanding single channel path, this session possesses a more complicated propagation path arrivals from two outstanding channel paths, and sometimes they interact with each other. This makes the interpolation more difficult and less accurate.

Results presented in Fig. 11 demonstrate that the receiver using all the three beamforming techniques show improved detection

performance at both spectral efficiencies with the SNR increasing from −14 dB to 14 dB compared to that without using beamforming technique. The receiver using the TFT-CPSD beamforming technique provides better performance than both the other beamforming techniques at both the spectral efficiencies of 1 bps/Hz and 1/2 bps/Hz. At a lower spectral efficiency (1/2 bps/Hz), the receiver using the TFT-CPSD beamforming technique achieves error-free transmission at SNR ≥ 7 dB, while the receiver using the other two beamforming techniques is unable to achieve error-free transmission at such SNR of 7 dB. This illustrates that the receiver with the FD beamforming and the TFT-CSDM beamforming is inferior to process such complex case of channel arrivals from multiple interacted directions than that of using the TFT-CPSD beamforming technique in UWA channels. The TFT-CPSD performs better than the other two beamforming techniques due to its reduced energy leakage with overlapped frames and its fully computed energy of received signals.

6. Verification of beamforming leakage

To verify the merit of no beamforming leakage from the proposed TFT-CPSD beamforming, we use the Waymark propagation model based simulation (Liu et al., 2012; Henson et al., 2014; Li et al., 2018a). In the simulation, the transmitter is stationary at a depth of 300 m. The receive VLA is towed by an ocean surface platform, and has a periodic oscillation with a maximum oscillating angle of $\vartheta_M = 1.5°$, as shown in Fig. 4. When the oscillating angle $\vartheta(t) = 0°$, the depth of the first hydrophone is 300 m, and the distance between the transmitter and the receive VLA is 60 km. The SSP used in the simulation is shown in Fig. 5.

During the simulation, 200 guard-free OFDM symbols are continuously transmitted. The receive VLA oscillation is considered to be induced by the sea current/turbulence, consistent with that from the two

(a) 1 bps/Hz.

(b) 1/2 bps/Hz; error-free transmissions are not shown.
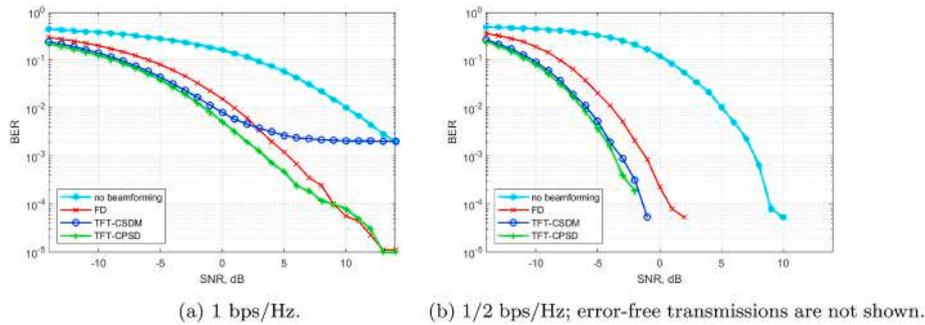
Fig. 10. BER performance of the receiver without using beamforming and with the DOA estimator using the three beamforming techniques in the F1-1 session as a function of SNR at different spectral efficiencies. (a) 1 bps/Hz (1024 bits/s); (b) 1/2 bps/Hz (512 bits/s).

(a) 1 bps/Hz.

(b) 1/2 bps/Hz; error-free transmissions are not shown.
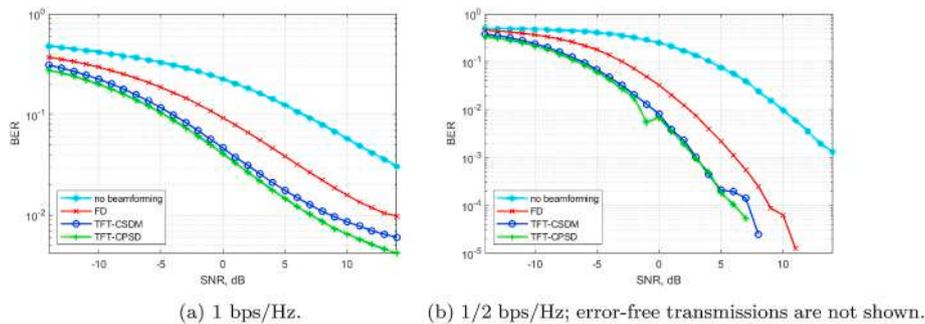
Fig. 11. BER performance of the receiver without using beamforming and with the DOA estimator using the three beamforming techniques in the F-3 session as a function of SNR at different spectral efficiencies. (a) 1 bps/Hz (1024 bits/s); (b) 1/2 bps/Hz (512 bits/s).
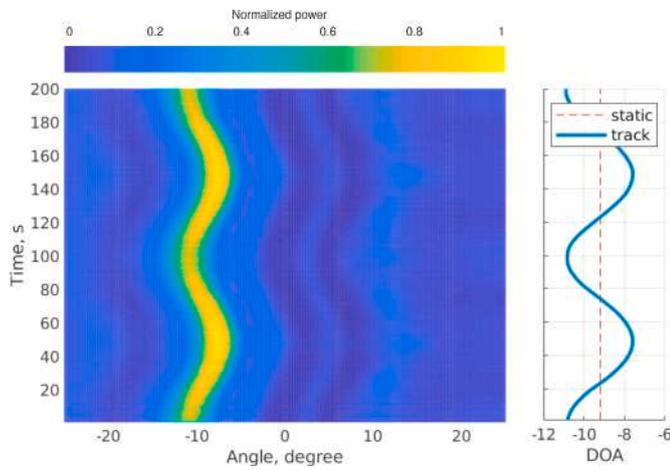
Fig. 12. DOA fluctuation in the simulation. Transmitter to receive VLA distance is 60 km. Left: estimated spatial signal power; right: angle of spatial power peak for the entire session (red dashed line), and angle of spatial power peak for each data frame (blue solid line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)
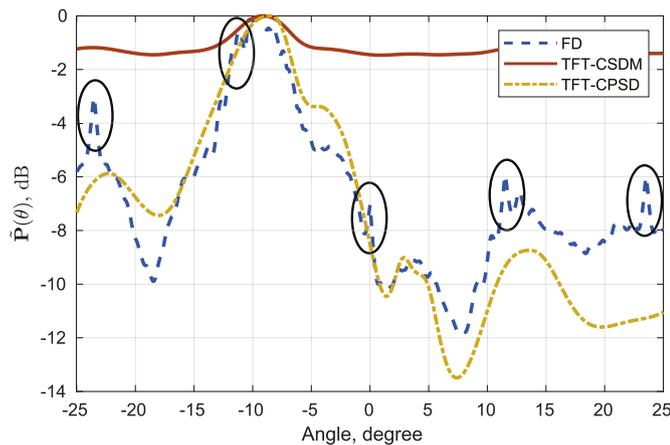
Fig. 13. Comparisons of average spatial signal power $\tilde{P}(\theta)$ estimated from the DOA estimator using the three beamforming techniques with data from the simulation without adding noise in the channel. Beamforming leakage from the FD beamforming technique has been indicated in black circles, while the two TFT beamforming do not have such periodical leaking peaks.

sea trial sessions as shown in Fig. 6. The oscillating angle is given by

$$\vartheta(t) = -\vartheta_M \cos\left(\frac{2\pi t}{T_p}\right), t \in [0, T-1], \tag{26}$$

where $T_p = 100$ s is the period of the VLA oscillation, and $T = 200$ s the duration of the communication session, ignoring propagation time in the channel. Note that when the angle is on the left hand of the middle dashed vertical line (see Fig. 4), the $\vartheta(t)$ is set as a negative value; and vice versa.

Fig. 12 left side shows spatial power distribution in the simulation. It can be seen that an outstanding cluster is identified as the one with DOA around $\hat{\theta} = -9.2°$. Fig. 12 right side shows the time-varying DOA crossing the static DOA $\hat{\theta} = -9.2°$ computed from the average spatial signal power for the entire communication session.

Fig. 13 shows comparison results of average spatial signal power $\tilde{P}(\theta)$ estimated from the DOA estimator using the three beamforming

techniques with data from the Waymark model simulation without adding channel noise. The DOA estimator using the proposed TFT-CPSD beamforming outperforms that of using the TFT-CSDM beamforming in accuracy, while it is comparable to that using the FD beamforming. In this case of without adding channel noise, there are still multiple extra power peaks (black circles in Fig. 13) from the result of FD beamforming, which is the same as that with the sea trial data shown in Fig. 3. The result indicates that these peaks are from beamforming leakage associated with the FD beamforming rather than from the underwater ambient. Such leakage of the target signal into the interference at multiple angles is interpreted from the time domain interpolation, and makes the DOA detection difficult, especially at low SNR. The TFT-CPSD beamforming does not have such beamforming leakage problem while provides high detection accuracy.

7. Conclusions and discussion

In this paper, we exploit the capability of three beamforming techniques, including a proposed low complexity TFT-CPSD (time-frequency-time with cross power spectral density) beamforming, in time-varying underwater acoustic communication channels, for improving the accuracy of DOA estimation and the detection performance of a receive system. The investigated receiver is designed for an underwater acoustic communication system with the transmission of guard-free OFDM signals with superimposed pilot symbols. Sea trial results demonstrate that the DOA estimator using the proposed TFT-CPSD beamforming possesses higher accuracy than that of using an existing TFT-CSDM beamforming and lower complexity than that of using interpolation based FD beamforming. The receiver using the TFT-CPSD beamforming based DOA estimator outperforms that of using the FD beamforming and the TFT-CSDM beamforming in both relatively simple and complex underwater acoustic communication channels. Further, we verified low beamforming leakage from the proposed TFT-CPSD method. As the proposed beamforming technique is based on the investigation of energy conservative (better than the TFT-CSDM beamforming) and energy leakage reduction (better than the FD beamforming), which is not relative to modulation schemes, thus it is also available to be applied and tested with other modulation schemes.

As the curvature of wave-front in shallow water is much more complicated than deep water transmission due to multipath and the gradient of sound, the channel can sometimes even be considered as sparse. In such case we may not be able to find a specific direction of arrival (DOA). To solve such a more complicated problem, a technique considering both the proposed TFT-CPSD beamforming as well as an adaptive sparse filter (Berger et al., 2009; Tang et al., 2011) may need to be investigated in the following works. Here we consider one DOA for each experiment session, while there might be multiple arrivals from different directions, then we need to consider a combining technique, e. g. maximum ratio combining, and an adaptive, e.g. angle-dependent, Doppler estimation technique. For such two techniques, readers are referred to the literature (Li and Zakharov, 2018) and (Li et al., 2018a). The frame length investigated here may be specific for the experimental data collected in the northwest Pacific Ocean at a specific sea state. However, for the using of such proposed method, we suggest an investigation of the frame length with a test channel data prior to the application of it. Furthermore, as we can see from Fig. 6, the DOA is not constant through the entire session and can experience a fluctuation of up to 3°. Considering such fluctuated DOA as a constant DOA may be an inferior way than fully tracking the actual DOA. As we can reduce the energy leakage by applying a proper way, either using the overlap-save or overlap-add or convolution methods, we expect that a DOA tracking algorithm considering the energy peak for each frame can be developed to improve the SNR and receiver performance.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Ambient noises correlation among hydrophone channels

Ambient noises on different acoustic sensor channels have often been assumed as uncorrelated and Gaussian distributed in UWA communications (Rafati et al., 2013; Pelekanakis and Baggeroer, 2011; Song et al., 2011; van Walree et al., 2017; Cho et al., 2011; Roy et al., 2007; Kaddouri et al., 2013), which is a simplified process of noise in real ocean scenarios. Kilfoyle et al. (2005) pointed out that such simplification may significantly change the capacity value of channel spatial modulation. Here we present the cross-correlation of underwater ambient noise based on sea trial data measured by the vertical linear array (VLA) of different hydrophone channels to provide ocean acoustician as an instruction on this issue.

To show the strength of linear relationship between two variables, the Pearson correlation coefficient (Benesty et al., 2009) is computed as

$$\xi = \frac{\sum_{k=1}^{K}(\phi_1(k) - \overline{\phi}_1)(\phi_2(k) - \overline{\phi}_2)}{\sqrt{\varepsilon_1^2}\sqrt{\varepsilon_2^2}}, \tag{A.1}$$

where

$$\varepsilon_1^2 = \sum_{k=1}^{K}\left(\phi_1(k) - \overline{\phi}_1\right)^2, \tag{A.2}$$

and

$$\varepsilon_2^2 = \sum_{k=1}^{K}\left(\phi_2(k) - \overline{\phi}_2\right)^2, \tag{A.3}$$

are covariance of the variables, $\phi_1(k)$ and $\phi_2(k)$ are the two variables, $\overline{\phi}_1$ and $\overline{\phi}_2$ are mean values of the two variables, respectively. Values between 0 and 0.3 (0 and −0.3) indicate a weak positive (negative) linear relationship via a shaky linear rule; values between 0.3 and 0.7 (−0.3 and −0.7) indicate a moderate positive (negative) linear relationship via a fuzzy-firm linear rule; and values between 0.7 and 1.0 (−0.7 and −1.0) indicate a strong positive (negative) linear relationship via a firm linear rule (Benesty et al., 2009).

(a) Ambient noise measured by the VLA in the sea trial   (b) Randomly distributed white Gaussian noise
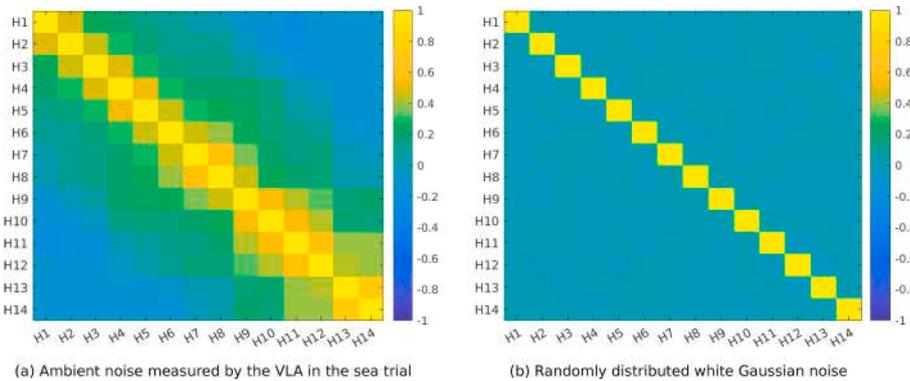
Fig. A14. Pearson correlation coefficient between different hydrophone channel noise in three cases. (a) Measured ambient noises in the sea trial; (b) White Gaussian noises. Negative correlations are in blue and positive correlations in yellow. 'H·' represents hydrophone index.

Before the sea trial communication sessions, we measured the ambient noise for each hydrophone channel at its depth (420 m + hydrophone distances [0 0.6 1.2 1.8 2.4 3.0 3.6 3.9 4.8 5.4 6.0 6.6 7.8 8.1] m) using the receive VLA. Fig. A14(a) shows the overall Pearson correlation coefficients computed from (A.1) with 20 s measured ambient noise for all the 14 hydrophone channels. As a result, the positive correlation between ambient noises measured by two hydrophones gradually decreases from strong to weak as the distance between the two hydrophones increases. Ambient noises measured by different hydrophone channels show strong/moderate positive correlation from two neighbour hydrophones with distance less than 0.6 m, show week correlation when the distance between two hydrophones is from 0.6 m to 4.0 m, and can only be considered as uncorrected where and when the distance between two hydrophones is more than 4.0 m. For comparison, we also show the Pearson correlation coefficient of randomly distributed white Gaussian noise in Fig. A14(b), which indicates uncorrected relationships between them.

As can be seen from Fig. A14, when we add noise on the received signals for each hydrophone channel to obtain signals with target SNR for beamforming, we need to consider the specific measurement depth, position, and ocean bathymetry related correlation of channel ambient noise, to ensure correct capacity value of channel spatial modulation is obtained. Such specific correlation between channel noise makes influence on the beamforming performance. Detailed influence is out of the scope in this paper, and will be a research topic in a following work.

References

Alexander, S.T., 1986. Adaptive Signal Processing: Theory and Applications. Springer-Verlag New York, Inc.

Asolkar, P., Das, A., Gajre, S., Joshi, Y., 2017. Comprehensive correlation of ocean ambient noise with sea surface parameters. Ocean Eng. 138, 170–178. https://doi.org/10.1016/j.oceaneng.2017.04.033.

Bai, Y., Li, J., Wu, Y., Wang, Q., Zhang, X., 2019. Weighted incoherent signal subspace method for DOA estimation on wideband colored signals. IEEE Access 7, 1224–1233. https://doi.org/10.1109/ACCESS.2018.2886250.

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson correlation coefficient. In: Noise Reduction in Speech Processing. Springer, pp. 1–4.

Berger, C.R., Zhou, S., Preisig, J.C., Willett, P., 2009. Sparse channel estimation for multicarrier underwater acoustic communication: from subspace methods to compressed sensing. IEEE Trans. Signal Process. 58 (3), 1708–1721. https://doi.org/10.1109/TSP.2009.2038424.

Brooker, A., Humphrey, V., 2016. Measurement of radiated underwater noise from a small research vessel in shallow water. Ocean Eng. 120, 182–189. https://doi.org/10.1016/j.oceaneng.2015.09.048.

Capon, J., 1969. High-resolution frequency-wavenumber spectrum analysis. Proc. IEEE 57 (8), 1408–1418. https://doi.org/10.1109/PROC.1969.7278.

Cho, S.E., Song, H.C., Hodgkiss, W.S., 2011. Successive interference cancellation for underwater acoustic communications. IEEE J. Ocean. Eng. 36 (4), 490–501. https://doi.org/10.1109/JOE.2011.2158014.

Daher, A., Baghious, E.H., Burel, G., Radoi, E., 2010. Overlap-save and overlap-add filters: optimal design and comparison. IEEE Trans. Signal Process. 58 (6), 3066–3075. https://doi.org/10.1109/TSP.2010.2044260.

Haykin, S., Justice, J.H., Owsley, N.L., Yen, J.L., Kak, A.C., 1985. Array Signal Processing. Prentice-Hall, Inc., Englewood Cliffs, NJ.

Henson, B., Li, J., Zakharov, Y.V., Liu, C., 2014. Waymark baseband underwater acoustic propagation model. In: Underwater Communications and Networking (UComms). IEEE, pp. 1–5. https://doi.org/10.1109/UComms.2014.7017132.

Hodges, R.P., 2011. Underwater Acoustics: Analysis, Design and Performance of Sonar. John Wiley & Sons. https://doi.org/10.1002/9780470665244.

Hoshuyama, O., Sugiyama, A., Hirano, A., 1999. A robust adaptive beamformer with a blocking matrix using coefficient-constrained adaptive filters. IEICE Trans. Fund. Electron. Commun. Comput. Sci. 82 (4), 640–647.

Ijaz, S., Silva, A., Jesus, S.M., 2012. Arrival-based Equalizer for Underwater Communication Systems. UCOMMS'12, Sestri Levante, pp. 1–10. Italy.

Kaddouri, S., Beaujean, P.-P.J., Bouvet, P.-J., Real, G., 2013. Least square and trended Doppler estimation in fading channel for high-frequency underwater acoustic communications. IEEE J. Ocean. Eng. 39 (1), 179–188. https://doi.org/10.1109/JOE.2013.2282065.

Kellett, P., Turan, O., Incecik, A., 2013. A study of numerical ship underwater noise prediction. Ocean Eng. 66, 113–120. https://doi.org/10.1016/j.oceaneng.2013.04.006.

Kilfoyle, D.B., Preisig, J.C., Baggeroer, A.B., 2005. Spatial modulation experiments in the underwater acoustic channel. IEEE J. Ocean. Eng. 30 (2), 406–415. https://doi.org/10.1109/JOE.2004.834168.

Li, J., Zakharov, Y.V., 2018. Efficient use of space-time clustering for underwater acoustic communications. IEEE J. Ocean. Eng. 43 (1), 173–183. https://doi.org/10.1109/JOE.2017.2688558.

Li, J., Stoica, P., Wang, Z., 2003. On robust Capon beamforming and diagonal loading. IEEE Trans. Signal Process. 51 (7), 1702–1715. https://doi.org/10.1109/TSP.2003.812831.

Li, J., Zakharov, Y.V., Henson, B., 2018a. Multibranch autocorrelation method for Doppler estimation in underwater acoustic channels. IEEE J. Ocean. Eng. 43 (4), 1099–1113. https://doi.org/10.1109/JOE.2017.2761478.

Li, D.-Q., Hallander, J., Johansson, T., 2018b. Predicting underwater radiated noise of a full scale ship with model testing and numerical methods. Ocean Eng. 161, 121–135. https://doi.org/10.1016/j.oceaneng.2018.03.027.

Li, J., White, P.R., Bull, J.M., Leighton, T.G., 2019a. A noise impact assessment model for passive acoustic measurements of seabed gas fluxes. Ocean Eng. 183 (1), 294–304. https://doi.org/10.1016/j.oceaneng.2019.03.046.

Li, J., White, P.R., Roche, B., Davis, J.W., Leighton, T.G., 2019b. Underwater radiated noise from hydrofoils in coastal water. J. Acoust. Soc. Am. 146 (5), 3552–3561. https://doi.org/10.1121/1.5134779.

Li, B., Zheng, S., Tong, F., 2019c. Bit-error rate based Doppler estimation for shallow water acoustic OFDM communication. Ocean Eng. 182, 203–210. https://doi.org/10.1016/j.oceaneng.2019.04.045.

Li, J., White, P.R., Roche, B., 2020. Seafloor noise ensemble from vessel manoeuvre in the central North Sea. Ocean Eng. 196, 106836. https://doi.org/10.1016/j.oceaneng.2019.106836.

Liu, J., Lin, I., Chu, C., 2005. Effects of sediment properties on surface-generated ambient noise in a shallow ocean. Ocean Eng. 32 (16), 1887–1905. https://doi.org/10.1016/j.oceaneng.2005.04.001.

Liu, C., Zakharov, Y.V., Chen, T., 2012. Doubly selective underwater acoustic channel model for a moving transmitter/receiver. IEEE Trans. Veh. Technol. 61 (3), 938–950. https://doi.org/10.1109/TVT.2012.2187226.

Ma, L., Dickson, K., McAllister, J., McCanny, J., 2011. QR decomposition-based matrix inversion for high performance embedded MIMO receivers. IEEE Trans. Signal Process. 59 (4), 1858–1867. https://doi.org/10.1109/TSP.2011.2105485.

Muramatsu, S., Kiya, H., 1997. Extended overlap-add and-save methods for multirate signal processing. IEEE Trans. Signal Process. 45 (9), 2376–2380. https://doi.org/10.1109/78.622960.

Oppenheim, A.V., Buck, J.R., Schafer, R.W., 2001. In: Discrete-time Signal Processing, vol. 2. Prentice Hall, Upper Saddle River, NJ.

Ozanich, E., Gerstoft, P., Niu, H., 2020. A feedforward neural network for direction-of-arrival estimation. J. Acoust. Soc. Am. 147 (3), 2035–2048. https://doi.org/10.1121/10.0000944.

Pelekanakis, K., Baggeroer, A.B., 2011. Exploiting space–time–frequency diversity with MIMO–OFDM for underwater acoustic communications. IEEE J. Ocean. Eng. 36 (4), 502–513. https://doi.org/10.1109/JOE.2011.2165758.

Pizzuti, L., dos Santos Guimarães, C., Iocca, E.G., de Carvalho, P.H.S., Martins, C.A., 2012. Continuous analysis of the acoustic marine noise: a graphic language approach. Ocean Eng. 49, 56–65. https://doi.org/10.1016/j.oceaneng.2012.04.004.

Proakis, J.G., 1995. Digital Communications. McGraw-Hill. New York.

Quinquis, A., Boulinguez, D., 1997. Multipath channel identification with wavelet packets. IEEE J. Ocean. Eng. 22 (2), 342–346. https://doi.org/10.1109/48.585953.

Rafati, A., Lou, H., Xiao, C., 2013. Soft-decision feedback turbo equalization for LDPC-coded MIMO underwater acoustic communications. IEEE J. Ocean. Eng. 39 (1), 90–99. https://doi.org/10.1109/JOE.2013.2241933.

Raghunath, K.J., Reddy, U.V., 1992. Finite data performance analysis of MVDR beamformer with and without spatial smoothing. IEEE Trans. Signal Process. 40 (11), 2726–2736. https://doi.org/10.1109/78.165659.

Roy, S., Duman, T.M., McDonald, V., Proakis, J.G., 2007. High-rate communication for underwater acoustic channels using multiple transmitters and space–time coding: receiver structures and experimental results. IEEE J. Ocean. Eng. 32 (3), 663–688. https://doi.org/10.1109/JOE.2007.899275.

Song, A., Abdi, A., Badiey, M., Hursky, P., 2011. Experimental demonstration of underwater acoustic communication by vector sensors. IEEE J. Ocean. Eng. 36 (3), 454–461. https://doi.org/10.1109/JOE.2011.2133050.

Sutton, J.L., 1979. Underwater acoustic imaging. Proc. IEEE 67 (4), 554–566. https://doi.org/10.1109/PROC.1979.11283.

Swingler, D.N., 1999. A low-complexity MVDR beamformer for use with short observation times. IEEE Trans. Signal Process. 47 (4), 1154–1160. https://doi.org/10.1109/78.752616.

Tang, Z., Blacquiere, G., Leus, G., 2011. Aliasing-free wideband beamforming using sparse signal representation. IEEE Trans. Signal Process. 59 (7), 3464–3469. https://doi.org/10.1109/TSP.2011.2140108.

Vaidyanathan, C., Buckley, K.M., 1995. Performance analysis of the MVDR spatial spectrum estimator. IEEE Trans. Signal Process. 43 (6), 1427–1437. https://doi.org/10.1109/78.388855.

van Walree, P.A., Socheleau, F.-X., Otnes, R., Jenserud, T., 2017. The watermark benchmark for underwater acoustic modulation schemes. IEEE J. Ocean. Eng. 42 (4), 1007–1018. https://doi.org/10.1109/JOE.2017.2699078.

Wang, Z., Zhou, S., Giannakis, G.B., Berger, C.R., Huang, J., 2011. Frequency-domain oversampling for zero-padded OFDM in underwater acoustic communications. IEEE J. Ocean. Eng. 37 (1), 14–24. https://doi.org/10.1109/JOE.2011.2174070.

Welch, P., 1967. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans. Audio Electroacoust. 15 (2), 70–73.

Wittekind, D., Schuster, M., 2016. Propeller cavitation noise and background noise in the sea. Ocean Eng. 120, 116–121. https://doi.org/10.1016/j.oceaneng.2015.12.060.

Xu, Z., Zakharov, Y.V., Kodanev, V.P., 2007. In: Space-time Signal Processing of OFDM Signals in Fast-Varying Underwater Acoustic Channel. IEEE OCEANS 2007-Aberdeen, pp. 1–6. https://doi.org/10.1109/OCEANSE.2007.4302328. UK.

Xu, Z., Li, H., Yang, K., 2020. A modified differential beamforming and its application for DOA estimation of low frequency underwater signal. IEEE Sensor. J. 1. https://doi.org/10.1109/JSEN.2020.2988025.

Yang, T., 2007. A study of spatial processing gain in underwater acoustic communications. IEEE J. Ocean. Eng. 32 (3), 689–709. https://doi.org/10.1109/JOE.2007.897072.

Zakharov, Y.V., Kodanev, V.P., 1994. Experimental study of an underwater acoustic communication system with pseudonoise signals. Acoust Phys. 40 (5), 707–715.

Zakharov, Y.V., Morozov, A.K., 2015. OFDM transmission without guard interval in fast-varying underwater acoustic channels. IEEE J. Ocean. Eng. 40 (1), 144–158. https://doi.org/10.1109/JOE.2013.2296842.

Zhao, Y., Yu, H., Wei, G., Ji, F., Chen, F., 2016. Parameter estimation of wideband underwater acoustic multipath channels based on fractional Fourier transform. IEEE Trans. Signal Process. 64 (20), 5396–5408. https://doi.org/10.1109/TSP.2016.2582466.

Zhou, S., Wang, Z., 2014. OFDM for Underwater Acoustic Communications. John Wiley & Sons.