Bountourakis, Vasileios; McCormack, Leo; Winberg, Mathias; Pulkki, Ville

Parametric spatial post-filtering utilising high-order circular harmonics with applications to underwater sound-field visualisation

Published in:
The Journal of the Acoustical Society of America

DOI:
10.1121/10.0005414

Published: 24/06/2021

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Bountourakis, V., McCormack, L., Winberg, M., & Pulkki, V. (2021). Parametric spatial post-filtering utilising high-order circular harmonics with applications to underwater sound-field visualisation. *The Journal of the Acoustical Society of America*, *149*(6), 4463-4476. https://doi.org/10.1121/10.0005414

Parametric spatial post-filtering utilising high-order circular harmonics with applications to underwater sound-field visualisation

Vasileios Bountourakis, Leo McCormack, Mathias Winberg, and Ville Pulkki

Citation: The Journal of the Acoustical Society of America **149**, 4463 (2021); doi: 10.1121/10.0005414
View online: https://doi.org/10.1121/10.0005414
View Table of Contents: https://asa.scitation.org/toc/jas/149/6
Published by the Acoustical Society of America

JASA ARTICLE

Parametric spatial post-filtering utilising high-order circular harmonics with applications to underwater sound-field visualisation

Vasileios Bountourakis,[1,a)] Leo McCormack,[1] Mathias Winberg,[2] and Ville Pulkki[1,b)]
[1]*Department of Signal Processing and Acoustics, Aalto University, Espoo 02150, Finland*
[2]*Saab Kockums AB, Malmö 211 19, Sweden*

ABSTRACT:
Beamforming using a circular array of hydrophones may be employed for the task of two-dimensional (2D) underwater sound-field visualisation. In this article, a parametric spatial post-filtering method is proposed, which is specifically intended for applications involving large circular arrays and aims to improve the spatial selectivity of traditional beamformers. In essence, the proposed method is a reformulation of the cross-pattern coherence (CroPaC) spatial post-filter, which involves calculating the normalised cross-spectral density between two signals originating from coincident beamformers. The resulting parameter may be used to sharpen another beamformer steered in the same look-direction, while attenuating ambient noise and interferers from other directions. However, while the original 2D version of the algorithm has been demonstrated to work well with second-order circular harmonic input, it becomes increasingly less suitable with increasing input order. Therefore, the proposed reformulation extends the applicability of CroPaC for much higher orders of circular harmonic input. The method is evaluated with simulated data of a 96-channel circular hydrophone array in three different passive sonar scenarios, where the proposed post-filter is shown to improve the spatial selectivity of both delay-and-sum and minimum-variance distortionless response beamformers. https://doi.org/10.1121/10.0005414
(Received 8 April 2021; revised 1 June 2021; accepted 2 June 2021; published online 24 June 2021)
[Editor: Jianlong Li] Pages: 4463–4476

I. INTRODUCTION

The subject of this article pertains to passive sonar systems, which can visualise the surrounding sound-field in order to aid submarine navigation or underwater surveillance. For this task, the spatial attributes of the underwater sound-field are captured by employing an array of hydrophones. Circular arrays are often well-suited for underwater navigation, as they permit full 360° sound-field imaging on the plane where the sensors are placed. To visualise the sound-field, a popular approach is to first steer a number of beamformers toward a dense grid of directions that sample a particular spatial area of interest. The energy of these beamformers may then be plotted against direction and time as a bearing-time record (BTR) (Brinkmann and Hurka, 2009), where bright spots imply the presence of sound emitting sources within the operating range of the sonar system. Therefore, a fundamental design optimisation for beamformer-based approaches is to minimise the widths of these beamformers, as this, in turn, improves the bearing accuracy of interesting targets.

The delay-and-sum (DaS) method is a popular example of a beamformer that has been widely used for such applications, likely owing to its simplicity, robustness to noise, and low computational complexity (Carter, 1981; Pirkl and Aughenbaugh, 2015; Quazi, 1981). More recently, beamformers formulated in the circular harmonic domain (CHD) have been proposed, which aim to achieve more consistent directivity characteristics over a wider frequency range when compared with the conventional DaS beamformer. These methods, also referred to as phase-mode beamformers, have been mainly studied and tested in the air domain with circular microphone arrays (Meyer, 2001; Parthy *et al.*, 2011; Tiana-Roig *et al.*, 2010), but also to a lesser extent in the context of underwater array processing (Liu, 2012). Typically, the spatial selectivity of these linear beamforming techniques is predominantly influenced by the array radius and the number of sensors (Van Trees, 2004). Signal-dependent beamformers, on the other hand, can achieve higher resolution for the same array geometry but can be less robust to ambient noise and coherent interferers.

In this article, a parametric spatial post-filtering method is proposed for the task of improving the spatial selectivity of existing beamformers. The proposed method is a novel reformulation of the cross-pattern coherence (CroPaC) post-filter (Delikaris-Manias and Pulkki, 2013), which was originally designed for compact microphone arrays utilising up to second-order circular (or spherical) harmonic input and has been shown to be especially suited to sound-field visualisation in noisy and reverberant environments (McCormack *et al.*, 2017). The main motivation for this study was to

[a)]Electronic mail: vasileios.bountourakis@aalto.fi, ORCID: 0000-0001-6285-9839.
[b)]ORCID: 0000-0003-3460-9677.

adapt the CroPaC concept for use in the underwater domain and evaluate its performance in realistic passive sonar scenarios. Since submarines may employ arrays comprising tens to thousands of hydrophone sensors, the proposed formulation has been specifically tailored for these applications by taking full advantage of very high-order circular harmonics (CHs) as input. To the best of the authors' knowledge, there is no existing literature on CH processing in the underwater domain that considers arrays of more than 12 hydrophones (Liu, 2012).

This paper is organised as follows. Section II provides background regarding the DaS beamformer, used as the primary baseline approach in the evaluation section, and the theory regarding CHD processing, including spatially encoding hydrophone sensor signals into CH signals and computing the relevant beamforming weights. This is followed by the original formulation of the two-dimensional (2D) CroPaC algorithm. The proposed algorithm is then presented in Sec. III and is evaluated using simulated data from a 96-channel circular array in Sec. IV. Finally, Sec. V discusses conclusions and potential future research opportunities based on the present work.

II. BACKGROUND

A. Underwater sound-field visualisation techniques

Traditionally, underwater sound-field visualisation involves techniques based on determining the energy of beamformers over a dense scanning grid. The relative energy of these beamformers is then subsequently depicted as a BTR and presented to the navigator of the marine vessel. Perhaps the most prolific beamforming algorithm is the DaS method (Carter, 1981; Quazi, 1981). The main reasons for its popularity pertain to its simplicity, robustness to noise, and low computational requirements. The method relies solely on applying appropriate time delays and gain factors for each hydrophone channel, followed by summation, all of which may also be realised with analogue circuits. Another category of linear and signal-independent beamforming techniques involves phase-mode beamformers, which are formulated in either the circular or the spherical harmonic domain. Circular harmonic beamforming (CHB) has been tested with circular microphone arrays mounted onto spherical (Meyer, 2001) or cylindrical (Parthy *et al.*, 2011; Tiana-Roig *et al.*, 2010) baffles and has been shown to yield improved and more consistent performance, compared to DaS beamforming, in terms of directivity index (DI), white noise gain, and sidelobe characteristics. A similar comparison for the case of spherical microphone arrays was performed in Rafaely (2005). Moreover, CH beamformers offer the advantage of implementation flexibility, since the beamforming stage is decoupled from the encoding stage, i.e., the sensor positions do not need to be known by the beamformer algorithm (Yan, 2020). Teutsch and Kellermann (2006) also reformulated traditional source detection algorithms from the space domain to the CHD and obtained a significant advantage in terms of detecting multiple simultaneous wideband sources. In Torres *et al.* (2012), the CH framework is employed in combination with time-frequency processing for robust source localisation in noisy and reverberant environments. The potential of CH beamforming for underwater acoustic systems has also been explored in Liu (2012), McIntyre *et al.* (2015), and Zou and Nehorai (2009).

The general downside of linear beamforming approaches, however, is that their spatial resolution is inherently limited by the number of hydrophones or, equivalently, the maximum order of CH expansion. Therefore, signal-dependent alternatives may be employed for this task. One popular example is the minimum-variance distortionless response (MVDR) beamformer (Ferguson and Carevic, 2010; Van Veen and Buckley, 1988; Zoltowski, 1988), which aims to minimise the beamformer signal energy while constraining the pattern to unity in the look-direction. Essentially, this approach implicitly steers the nulls of the beamformer pattern toward interfering sources, thus improving the resolution of the BTR image. If the directions of the interfering sources are known or can be estimated, then the nulls may be placed explicitly through a linearly constrained minimum-variance (LCMV) solution (Frost, 1972; Habets *et al.*, 2009; Peled and Rafaely, 2011). In general, these adaptive approaches yield improved spatial selectivity over linear and time-invariant methods, with the penalty of increased computational requirements. Moreover, their performance degrades with increasing ambient noise, often reducing to that of conventional beamformers in very low signal-to-noise ratio (SNR) conditions (Zoltowski, 1988). The MVDR method is also known to be susceptible to coherent interferers (Peled and Rafaely, 2011; Reddy *et al.*, 1987), such as acoustical reflections from the water surface, and modeling errors, such as uncertainties in element positions and steering vectors (Cox, 1973).

Other signal-dependent approaches include subspace methods, such as the multiple signal classification (MUSIC) (Schmidt, 1986) or estimation of signal parameters via rotational invariance technique (ESPRIT) (Gao and Gershman, 2005; Jo *et al.*, 2020; Roy and Kailath, 1989) algorithms. Rather than determining beamformer energy, MUSIC operates by computing values that correspond to the likelihood of sound sources being located in each scanning direction. Plotting the resulting pseudo-spectrum can often yield higher resolution BTR images than beamformer-based approaches. ESPRIT, on the other hand, is a grid-less approach that can directly extract sound source direction-of-arrival (DoA) estimates. The primary downside of subspace methods, however, is that they require source detection algorithms (Akaike, 1974; Chen *et al.*, 1991; Han and Nehorai, 2013) and can also be sensitive to correlated signals due to multipath propagation. Other grid-less DoA estimators include those derived from acoustic active-intensity (Fahy and Salmon, 1990; McCormack *et al.*, 2019; Moore *et al.*, 2017) and maximum likelihood estimators (Bianchi *et al.*, 2015; Tervo and Politis, 2015).

An alternative approach, however, is to employ the use of spatial post-filters to improve upon the spatial selectivity

https://doi.org/10.1121/10.0005414

of existing beamformers and also deactivate them during periods of source inactivity. The application of spatial post-filters is therefore especially suited to environments with high background noise but has nonetheless received limited attention in the underwater acoustics community. Examples of such post-filters include multi-channel Wiener filters (Simmer *et al.*, 2001), the McCowan filter (McCowan and Bourlard, 2003), and CroPaC (Delikaris-Manias and Pulkki, 2013). Traditionally, these spatial post-filters have been used for audio signal enhancement applications. However, the CroPaC post-filter has also been employed for sound-field visualisation in McCormack *et al.* (2017), where it was shown to produce sharper images than MVDR beamformers and MUSIC in highly reverberant environments in the air domain.

B. Delay-and-sum (DaS) beamforming

DaS is a beamforming technique that applies appropriate time delays to the sensor signals to align them in time for a particular look-direction prior to summing them together. The beamformer output, when plotted for each look-direction, is proportional to the amplitude of incident plane waves from that direction. The method may be formulated in the time-frequency domain, whereby the input signals are first segmented into short overlapping time-windows and then decomposed into narrowband frequency bins, via either a short-time Fourier transform (STFT) or a bank of filters. The time delays then correspond to phase shifts.

Assuming an array of Q sensors, the time-frequency domain signals of the array are denoted with the vector $\mathbf{x}(k,i) \in \mathbb{C}^{Q\times 1}$, where k and i refer to the frequency and time indices, respectively. The output of the DaS beamformer in the steering direction $\mathbf{a}_l$ for the angular frequency ω_k is then given by

$$b_{\mathrm{DS}}(k,i,\mathbf{a}_l) = \frac{1}{Q}\mathbf{w}_{\mathrm{DS}}^{\mathrm{H}}(\omega_k,\mathbf{a}_l)\mathbf{x}(k,i), \tag{1}$$

where $\mathbf{w}_{\mathrm{DS}}(\omega_k,\mathbf{a}_l) \in \mathbb{C}^{Q\times 1}$ is the vector of the beamformer weights applied to each sensor, which are simply phase shifts corresponding to the steering direction.

C. Circular harmonic (CH) processing

The method proposed in this article is derived based upon a CH decomposition of the captured sound-field, where the sensor array signals are spatially encoded into signals with directional patterns that follow CHs of different orders and degrees.

1. Spatial encoding to circular harmonics (CHs)

Given a circular array of Q sensors at the positions $p_q = (r,\phi_q)$, $q = 1,\ldots,Q$, where r is the radius of the array and $\phi_q \in [-\pi,\pi)$ is the polar angle of the sensor q, spatial encoding refers to the conversion of the array signal vector $\mathbf{x}(t) \in \mathbb{R}^{Q\times 1}$ from the sensor domain to the CHD. This conversion relates to a projection of the sound-field (which is sampled at discrete points on the perimeter of the array) onto a set of CH basis functions to obtain a set of CH signals.

The CH basis functions employed in this study are real-valued and defined as (Parthy *et al.*, 2011)

$$C_{nm}(\phi) = \begin{cases} 1 & \text{if } n = 0 \\ \sqrt{2}\cos(n\phi) & \text{if } n > 0 \text{ and } m = n \\ \sqrt{2}\sin(n\phi) & \text{if } n > 0 \text{ and } m = -n, \end{cases} \tag{2}$$

where $\phi \in [-\pi,\pi)$ is the polar angle, and n refers to the order and $m \in [-n,n]$ to the degree of the CH functions. Spatial resolution increases as higher truncation orders of CHs are used for the sound-field approximation. An Nth order representation of the sound-field requires $2N+1$ CH components. The reader is referred to Zotter and Frank (2019) for further information on the properties of these functions.

Since spatial encoding is a frequency-dependent operation, the array signals are first transformed into the time-frequency domain. The vector $\mathbf{s}(k,i) \in \mathbb{C}^{(2N+1)\times 1}$ of the CH signals is then given by

$$\mathbf{s}(k,i) = \mathbf{W}(k)\mathbf{x}(k,i), \tag{3}$$

where $\mathbf{W}(k) \in \mathbb{C}^{(2N+1)\times Q}$ is the spatial encoding matrix. Note that the time and frequency indices are henceforth omitted for brevity of notation.

The spatial encoding matrix can be factorised into a frequency-independent matrix $\mathbf{E}_{\mathrm{DCHT}} \in \mathbb{R}^{(2N+1)\times Q}$, which performs the discrete circular harmonic transform (DCHT), and a frequency-dependent equalisation matrix $\mathbf{W}_{\mathrm{EQ}} \in \mathbb{C}^{(2N+1)\times(2N+1)}$ compensating for the baffle of the array and the directivity of the sensors

$$\mathbf{W} = \mathbf{W}_{\mathrm{EQ}}\mathbf{E}_{\mathrm{DCHT}}. \tag{4}$$

The DCHT matrix for uniform distribution of the sensors is given by

$$\mathbf{E}_{\mathrm{DCHT}} = \frac{1}{Q}\mathbf{C}^{\mathrm{T}}(\boldsymbol{\phi}_Q), \tag{5}$$

where $\mathbf{C}(\boldsymbol{\phi}_Q) \in \mathbb{R}^{Q\times(2N+1)}$ is a matrix containing the values of the CH functions for each sensor polar angle $\boldsymbol{\phi}_Q = [\phi_1,\ldots,\phi_Q]$.

The equalisation matrix is a diagonal matrix comprising order dependent weights as

$$\mathbf{W}_{\mathrm{EQ}} = \mathrm{diag}(w_0,w_1,w_1,w_2,w_2,\ldots,w_N,w_N), \tag{6}$$

with

$$w_n = \frac{1}{b_n}\frac{|b_n|^2}{|b_n|^2+\lambda^2}, \tag{7}$$

where b_n are the equalisation modal coefficients (Teutsch, 2007; Williams, 1999). The Tikhonov regularisation

parameter λ prevents the amplification of sensor noise (at low frequencies) from exceeding a specified threshold (Moreau *et al.*, 2006). For alternative regularisation methods, the reader is referred to Bernschütz *et al.* (2011) and Jin *et al.* (2014). For comprehensive objective measures to determine encoding performance, the reader is referred to Moreau *et al.* (2006) and Politis and Gamper (2017).

2. Beamforming in the circular harmonic domain (CHD)

CH beamforming (CHB) is a technique well-suited to circular arrays. Once the CH signals are obtained using Eq. (3), a CH beamformer may be realised via a matrixing operation as (Tiana-Roig *et al.*, 2010)

$$b_{\mathrm{CH}}(k,i) = \mathbf{w}_{\mathrm{CH}}^{\mathrm{T}}\mathbf{s}(k,i), \tag{8}$$

where $\mathbf{w}_{\mathrm{CH}} \in \mathbb{R}^{(2N+1)\times 1}$ is a vector of frequency-independent beamforming weights.

The beamforming weights for an arbitrary pattern $\mathbf{P} \in \mathbb{R}^{L\times 1}$, sampled at L points uniformly distributed on the circle, can be obtained by performing a DCHT on the pattern using the $\mathbf{E}_{\mathrm{DCHT}}$ matrix defined in Eq. (5) as

$$\mathbf{w}_{\mathrm{CH}} = \frac{1}{L}\mathbf{C}^{\mathrm{T}}(\boldsymbol{\phi}_L)\mathbf{P}, \tag{9}$$

with $\boldsymbol{\phi}_L = [\phi_1, \ldots, \phi_L]$. For an accurate synthesis of a pattern of order N_p, the number of the sampling points must be $L \geq 2N_p + 1$, and the order of the available CH components must be $N \geq N_p$. The generated pattern can then be steered to any angle ϕ_l, by also including an appropriate rotation matrix $\mathbf{R}(\phi_l) \in \mathbb{R}^{(2N+1)\times(2N+1)}$ (Ivanic and Ruedenberg, 1998),

$$b_{\mathrm{CH}}(k,i,\phi_l) = \mathbf{w}_{\mathrm{CH}}^{\mathrm{T}}\mathbf{R}(\phi_l)\mathbf{s}(k,i). \tag{10}$$

D. Cross-Pattern Coherence (CroPaC)

CroPaC is a post-filtering technique based on capturing the sound-field with two coincident beamformers, which have equal gain and the same polarity in the look-direction (Pulkki *et al.*, 2018). The normalised cross-spectral density between the signals of these beamformers, referred to as coherence in this context, is used as a post-filter in the time-frequency domain to modulate a third beamformer signal. The net result is an attenuation of signal energy originating from other directions and during periods of source inactivity. The main idea of the approach is that the coherence between the coincident-beamformer signals is maximum only when the beamformers are steered toward the DoA of the sound source. The method has been deployed both in the circular and in the spherical harmonic domain for sound-field visualisation and beamformer signal enhancement applications (Delikaris-Manias and Pulkki, 2014; Delikaris-Manias *et al.*, 2016; McCormack *et al.*, 2017).

The time domain signals of a circular array are initially transformed into the time-frequency domain and then encoded into the CHD, as described in Sec. II C 1. Each element of the resulting vector of CH signals in Eq. (3) corresponds to the signal that would be captured by a sensor that replicates the directional pattern of a CH function $C_{nm}(\phi)$. The computation of the post-filter requires four CH signals denoted by S_j for $j = 1, 2, 3, 4$ in the following. If $N \geq 2$ is the maximum order of available CH components, signals S_1 and S_2 are CH signals of orders N and $N-1$, respectively, and correspond to CH directional patterns that contain the cosine term ($m = n$) in Eq. (2). Signals S_3 and S_4 are the respective CH signals of orders N and $N-1$ whose directional patterns contain the sine term ($m = -n$). The normalised cross-spectral density G between the signals S_1 and S_2, when the beamformers are steered in direction ϕ_l, is estimated as

$$G(\phi_l,k,i) = \frac{2\Re[\Phi_{12}(\phi_l,k,i)]}{\sum_{j=1}^{4}\Phi_{jj}(\phi_l,k,i)}, \tag{11}$$

where $\Re$ denotes the real operator, $\Phi_{12}(\phi_l,k,i)$ refers to the cross-spectral density between signals S_1 and S_2, and $\Phi_{jj}(\phi_l,k,i)$ are the auto-power spectral densities of signals S_j for $j = 1, 2, 3, 4$. These may be computed as

$$\Phi_{12}(\phi_l,k,i) = E\{S_1^*(\phi_l,k,i)S_2(\phi_l,k,i)\}, \tag{12}$$

$$\Phi_{jj}(\phi_l,k,i) = E\{|S_j(\phi_l,k,i)|^2\}, \tag{13}$$

where E denotes the expectation operator.

The denominator of Eq. (11) is a normalisation term, which equals twice the total energy of the sound-field. As a result, the calculated value of G lies within the range $[-1, 1]$. Negative values indicate that the captured signal originates from a direction where there is a mismatch in the phase of the beamformers. Since only waves arriving from the look-direction are of interest, the negative values may be truncated via rectification as

$$G_r(\phi_l,k,i) = \frac{(1+\beta)|G(\phi_l,k,i)| + (1-\beta)G(\phi_l,k,i)}{2}, \tag{14}$$

which, for $\beta = 0$, results in values of G_r within the range $[0, 1]$ (half-wave rectification).

Figure 1 shows the resulting G_r attenuation pattern when the cross-spectral density is calculated between CH signals of first and second order. In this case, signals S_1 and S_2 are obtained by a quadrupole and a dipole steered in the same look-direction, while signals S_3 and S_4 are obtained by the same patterns but rotated by 90°. The latter patterns are not depicted for graphical clarity. It can be observed that the highest value of G_r is obtained in the look-direction, where the two CH patterns have equal phase and maximum amplitude. Waves arriving from any other direction are attenuated, while there are also two back lobes of low amplitude at the angles where the two CH patterns have equal negative phase.

https://doi.org/10.1121/10.0005414

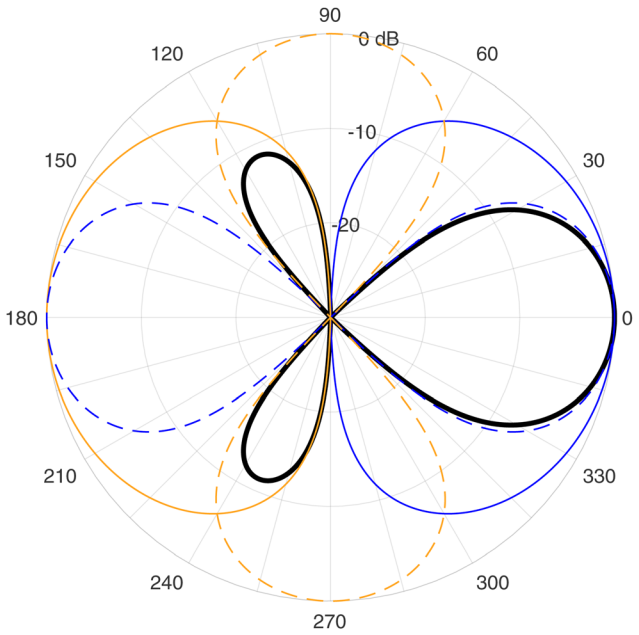


FIG. 1. (Color online) G_r attenuation pattern of 2D CroPaC (in bold). The pattern is computed based on the normalised cross-spectral density between two CH signals, which exhibit the directional patterns of a dipole (solid line) and a quadrupole (dashed line). The two directional patterns have maximum same-phase amplitude at 0°. The colours indicate phase information: blue for positive phase and orange for negative phase.

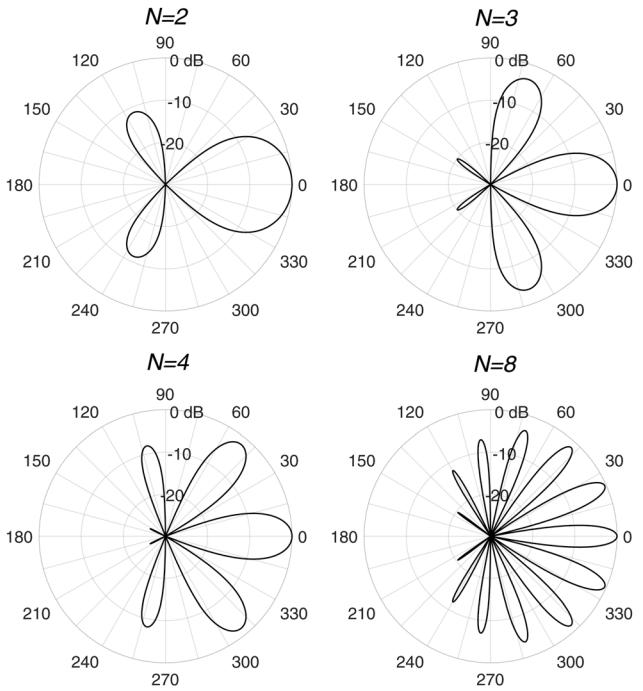


FIG. 2. G_r attenuation patterns of 2D CroPaC for different orders N of CH signals. The appearance of significant unwanted sidelobes is already evident from third order, which is where G_r is calculated using second and third order CH signals using the original formulation (Delikaris-Manias and Pulkki, 2013).

Finally, the signal $b(\phi_l, k, i)$ captured based on any other beamformer is then modulated by the G_r values over time and frequency to obtain the filtered signal as

$$y(\phi_l, k, i) = G_r(\phi_l, k, i) \cdot b(\phi_l, k, i). \tag{15}$$

III. PROPOSED METHOD FOR HIGHER ORDERS

A known issue with the original CroPaC algorithm (Delikaris-Manias and Pulkki, 2013) is that, with increasing input order, numerous unwanted sidelobes are introduced into the G_r directional attenuation pattern. Examples of these aberrations are depicted in Fig. 2. Note that the CroPaC "order" refers to the highest order N of CH signals used in the algorithm. It can be observed that the original 2D formulation (detailed in Sec. II D) becomes increasingly unsuitable above second order. This problem was first addressed by McCormack *et al.* (2017), who proposed a solution for the three-dimensional (3D) variant of CroPaC, which is afflicted with similar sidelobe characteristics in the spherical harmonic domain. The proposed solution involved iteratively rotating the CroPaC patterns along the look-direction axis, followed by multiplying them together. This process resulted in more directive CroPaC patterns but, more importantly, also greatly reduced the sidelobes, and thus, the use of higher-order spherical harmonic components was made feasible. However, a main downside of this solution is its reliance on a third spatial dimension to rotate the patterns. Therefore, this previous solution is not applicable to the 2D case, and thus, an alternative approach is proposed in this section specifically for 2D applications.

A. Processing principle

The new formulation of the CroPaC algorithm, henceforth referred to as higher-order cross-pattern coherence (HO-CroPaC), effectively solves the sidelobe problem of the original 2D CroPaC version, permitting the use of higher-order CH components. Its primary difference is in regard to the selection of the directional patterns of the signals employed for the calculation of G. Rather than using the CH directional patterns directly, patterns with more complex directional characteristics are first designed.

The problem of designing such directional patterns may be solved via a number of different approaches. In this work, a relatively simple heuristic approach is employed, which is demonstrated to perform well in Sec. IV. A set of four directional patterns are designed for capturing the signals S_j, $j = 1, 2, 3, 4$, needed for the computation of G, as in Eq. (11), with the following steps:

(a) The selection of two CH components of non-consecutive orders, and
(b) The multiplication of the CH patterns by a cardioid pattern steered in the look-direction.

These design choices for the directional patterns aim primarily at cancelling out the sidelobes of the G_r attenuation pattern. The first choice is based on the observation that, for high orders of expansion, the CH patterns of consecutive orders exhibit closely spaced sidelobes of the same phase. As a result, the multiplication of these patterns in the cross-spectral density calculation does not cancel out these

sidelobes as it does in the second-order case. In the original CroPaC version (Delikaris-Manias and Pulkki, 2013), the choice of utilising the two highest-order patterns was made to achieve maximum directivity. However, the method is still valid for any combination of two CH patterns of different orders due to the orthogonality property of the CH base functions. Therefore, by compromising on some of the directivity, the approach can instead prioritise the selection of patterns whose sidelobes appear at different angles or have opposite phase when they occur at those same angles, thereby leading to improved cancellation of the sidelobes.

The multiplication of the CH patterns by a cardioid pattern results in spatially weighted patterns toward the look-direction, which preserve the properties of the CH functions. By focusing the received energy in the look-direction, most of the remaining sidelobes of the previous step are further attenuated. This design choice may also improve the performance in cases where multiple sources in the sound-field exhibit energy content in the same time-frequency tile, a scenario that is rather common in underwater applications with multiple simultaneous broadband sources. It should be noted that the energy captured by these patterns, used to normalise the cross-spectral density [the denominator of Eq. (11)], corresponds to the local energy—in the spatial area covered by the cardioid pattern—and not to the total sound-field energy, as is the case for the original CroPaC formulation (Delikaris-Manias and Pulkki, 2013).

B. Proposed algorithm

In the original 2D CroPaC version, the signals S_j, $j = 1, 2, 3, 4$, used in Eqs. (12) and (13) are captured by the CH patterns given in Eq. (2). In the proposed method, those patterns are replaced by spatially weighted directional patterns, which are defined as follows:

$$\begin{aligned} P_1(\phi) &= P_{\text{card}}(\phi)\cos(N_{\text{high}}\phi),\\ P_2(\phi) &= P_{\text{card}}(\phi)\cos(N_{\text{low}}\phi),\\ P_3(\phi) &= P_{\text{card}}(\phi)\sin(N_{\text{high}}\phi),\\ P_4(\phi) &= P_{\text{card}}(\phi)\sin(N_{\text{low}}\phi), \end{aligned} \tag{16}$$

where N_{high} and N_{low} are the orders of the selected high and low order CHs, respectively, and $P_{\text{card}}(\phi)$ is a cardioid pattern of order N_{card} defined as

$$P_{\text{card}}(\phi) = \left(\frac{1}{2}\right)^{N_{\text{card}}}(1+\cos\phi)^{N_{\text{card}}}. \tag{17}$$

If the highest order of the available CH components is denoted by N, the desired beam patterns can be generated by beamforming in the CH domain, as described in Sec. II C 2, respecting the constraint $N_{\text{high}} + N_{\text{card}} \leq N$. The created patterns can then be steered in arbitrary look-directions along the horizontal plane by employing rotation matrices. The calculation of the G parameter, as well as the half-wave rectification and the post-filtering operations, are unaltered compared to the original formulation. A flow diagram of the process is depicted in Fig. 3.

An example of the resulting directional patterns $P_j(\phi)$, $j = 1, 2, 3, 4$ for $N_{card} = 10$, $N_{\text{high}} = 10$, and $N_{\text{low}} = 5$ ($N = 20$) is shown in Fig. 4. It can be observed that the multiplication of patterns P_1 and P_2 will result in a significant cancellation of their sidelobes, since their same-phase sidelobes occur in different directions. Figure 5 depicts the resulting G_r attenuation patterns obtained by (a) CroPaC for $N = 20$ and (b) HO-CroPaC utilising the directional patterns shown in Fig. 4. It can be observed that the proposed reformulation minimises the sidelobes at the expense of a wider mainlobe.

To further suppress the sidelobes, it is possible to also compute a second post-filtering value G'_r based on a different selection of directional patterns, such that the sidelobes of this second resulting pattern occur in different directions compared to the first one. The product of these two post-filtering values G_r and G'_r then results in the final post-filter, which offers reduced sidelobe behavior, increased resolution, and additional diffuse noise attenuation. Note that this dual post-filtering approach is employed henceforth for the evaluations.

IV. EVALUATION

A. Data description

The method was evaluated using simulated data of a 96-channel circular hydrophone array mounted onto a finite rigid cylindrical baffle. The radius of the array was $r = 1.35$ m, and the sampling frequency of the data was

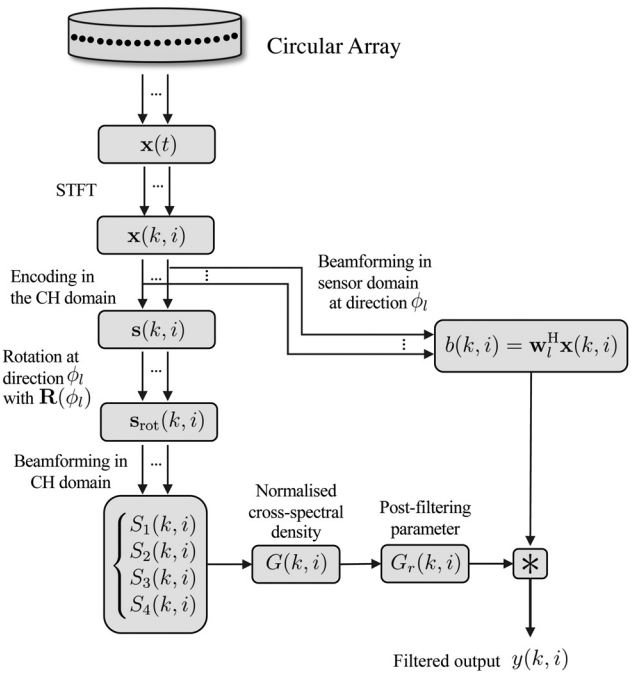


FIG. 3. Block diagram of the proposed HO-CroPaC filter. The process is repeated for each scanning grid direction ϕ_l. Note that any beamformer design may be used to obtain the baseline beamformer signal $b(k, i)$ in practice.

https://doi.org/10.1121/10.0005414

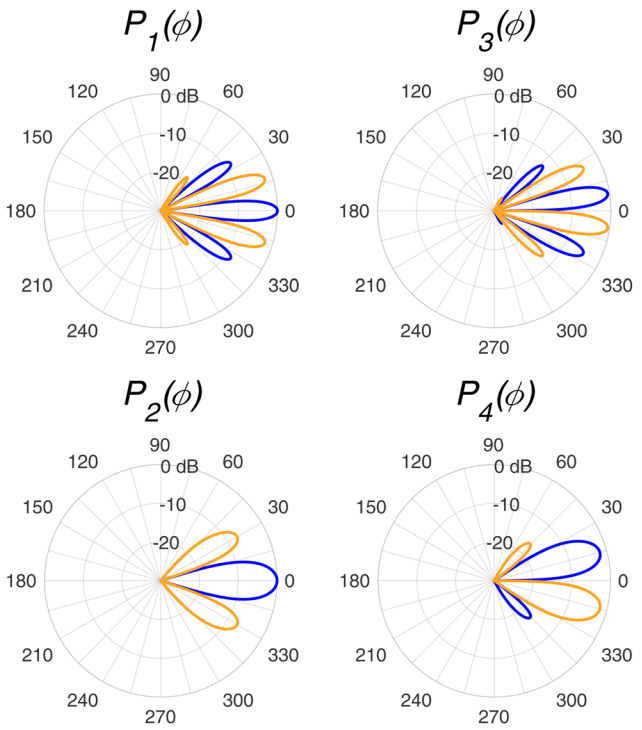


FIG. 4. (Color online) Directional patterns of signals used to compute attenuation values in the HO-CroPaC post-filter. Patterns P_1 and P_3 utilise CH components of order 20, while P_2 and P_4 utilise CH components of order 15. Patterns P_1 and P_2, used for the computation of the unnormalised cross-spectral density, are selected so that their same-phase sidelobes occur in different directions. All four patterns are used to determine the local sound-field energy, which is used to normalise the post-filter value.

$F_s = 31\,250$ Hz. The acoustic modeling employed for the data generation provides a highly detailed and sophisticated simulation of the properties of the targets (sound sources) and the underwater environment.

In the simulation model, the first step was to define the signal characteristics of the targets to be included within the simulation. Here, a model of a merchant vessel was employed, which was based on the combination of modulated broadband noise (to mimic propeller noise) and narrowband spectral energy (to model the engine noise). The second step was to simulate the underwater acoustical environment from the target to the submarine, including propagation loss, reverberation, and absorption. Additionally, ambient noise according to a certain sea state was added, as well as self-noise from the submarine housing the array. The environment parameters correspond to a Baltic Sea environment, which is a shallow water environment with a depth of typically around 40 m.

Three scenarios were evaluated. The first is a standard scenario often used for reference testing, while the other two are of particular interest in passive sonar applications.

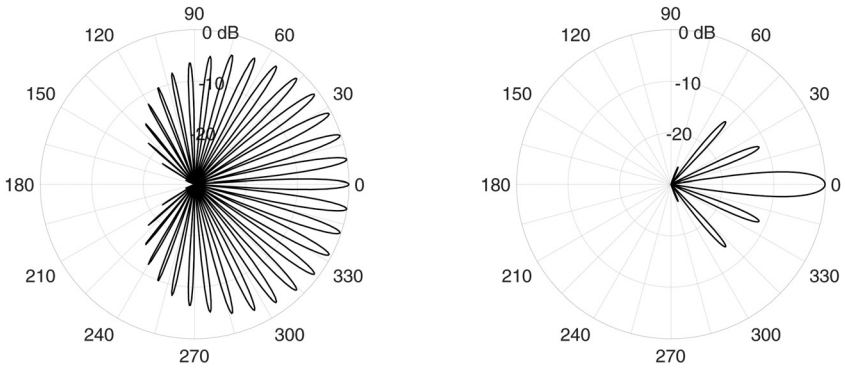
(1) Four moving sources in a diffuse field with low direct-to-diffuse ratio (DDR).
(2) A single moving source in shallow water, where a prominent reflection from the water surface is present.
(3) Two moving sources with bearings approaching one another, which are eventually in close proximity.

The generated data largely resemble the data captured with a real sonar system of similar number and arrangement of hydrophones for such underwater acoustic scenarios. For all the scenarios, the submarine to which the hydrophone array was attached was stationary with a heading of 0°; hence, the 0° angle always denotes the north direction, and no compensation for the relative movements of the submarine was required.

B. Implementation details

The conversion of the sensor signals into the time-frequency domain was performed via a STFT, utilising Hann windows of 100 ms with 50% overlap. Following this conversion, the signals were also bandpass-filtered between 5 and 9 kHz. The selection of this frequency range was based on preliminary observations of the frequency content of the sound sources in the studied scenarios. These frequencies are below the spatial aliasing limit of the array, with the simulations indicating that artefacts arising due to aliasing begin to occur above 10 kHz.

The DaS and MVDR beamformers were employed as baseline beamformers, upon which the proposed post-filter was applied. The scanning grid of their implementations



(a) CroPaC (b) HO-CroPaC

FIG. 5. Comparison of the G_r attenuation patterns obtained by CroPaC and HO-CroPaC with the same highest order of available CH signals. Pattern (a) is obtained with CH components of orders 19 and 20, while pattern (b) results from the patterns shown in Fig. 4.

consisted of 192 equally spaced directions on the horizontal plane, which corresponds to a spatial resolution of 1.875°. The DaS beamformer formulation given in Sec. II B served as the primary baseline beamformer design. For each generated beam, only half of the hydrophones (48) were employed, i.e., those on the side of the array toward the look-direction, which were also spatially filtered with a Hamming window. Note that this is a commonly utilised approach in practice for large arrays operating at low SNR levels. The MVDR beamformer was also implemented based on the formulation given in Yan (2015), which operates in the CH domain, and served as the secondary baseline beamformer. Note that the energy values of output beamformer signals were all temporally smoothed with a moving average filter over 15 successive time frames (averaging window of length 0.8 s). This temporal averaging was tuned to be short enough to adequately capture small changes in the target bearings while still providing suitably clear BTR images.

Regarding the implementation of HO-CroPaC, the encoding filters are derived using the theoretical CH decomposition (see Sec. II C 1), which assumes a circular array mounted onto an infinite cylindrical baffle. Teutsch and Kellermann (2006) showed that a finite cylinder whose length is 1.4 times its radius is enough to approximate an infinitely long cylinder. However, the baffle of the array under study is not high enough to satisfy this approximation. This inconsistency with the theoretical solution is expected to introduce a small error. The maximum order of CH components that can be obtained with the current array is $N = 47$. The higher-order directional patterns employed for the calculation of the post-filter are thus of order 47 (to maximise directivity). It should be noted that the higher the order of CH components used to create a beam-pattern, the more the low-frequency capsule noise is boosted due to the equalisation terms applied in the spatial encoding step. As a result, when higher-order harmonics are used for beamforming, the lower end of the frequency range is not usable due to contamination with sensor noise. The Tikhonov regularisation parameter λ, in Eq. (7), was set to 1.6×10^{-3} to prevent the amplification of noise in the frequency range of interest (5–9 kHz); thus, the highest-order harmonics could be used without issue.

Two post-filtering parameters are calculated with the following specifications:

- G_{r1}: $N_{\text{card}} = 20$, $N_{\text{high}} = 27$, $N_{\text{low}} = 13$,
- G_{r2}: $N_{\text{card}} = 34$, $N_{\text{high}} = 13$, $N_{\text{low}} = 6$.

The product of these two values $G_r = G_{r1} \times G_{r2}$ is used as the final post-filter parameter. The directional patterns employed for the calculation of G_{r1} are shown in Fig. 6. Note that this pattern selection is not unique and possibly not optimal. However, the authors deem this selection based on the heuristic approach described in Sec. III A as sufficient to demonstrate the capabilities of the method. The scanning grid and the time-frequency averaging settings are kept the same as used by the beamformer algorithms to facilitate comparison between the methods.

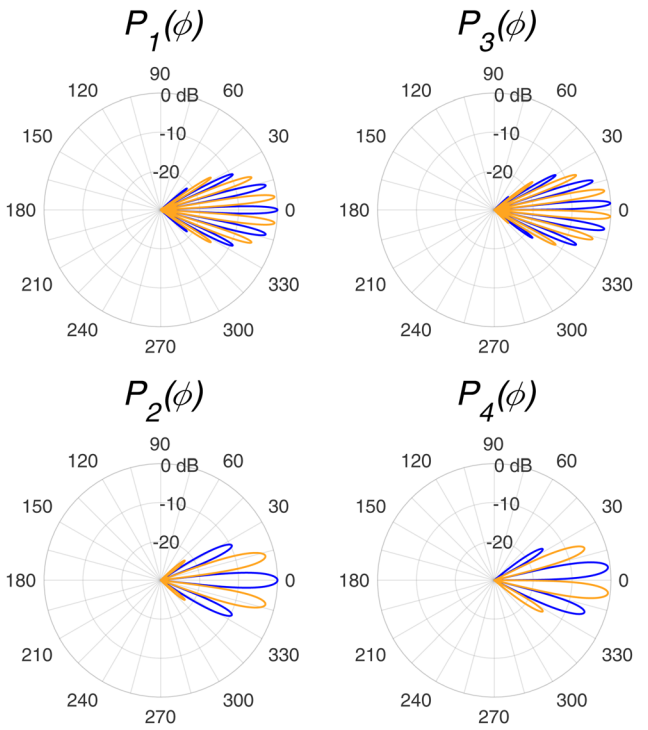


FIG. 6. (Color online) Directional patterns of signals employed for the calculation of G_{r1} utilising CH components of orders 47 (top) and 33 (bottom).

C. Results and discussion

In this section, the performance of the proposed method is investigated both numerically, by means of standard array performance measures, and visually, based on BTR plots. BTR plots depict the time evolution of the beamformer output energy at each steering angle (look-direction). The vertical axis is used for time and the horizontal axis for the bearings of the sound source(s). Time-frame slices of the BTR plots are also presented for better visualisation of the advantage obtained by applying the proposed post-filter. In the following plots, the BTR values are normalised to the maximum value obtained in each segment and presented in decibels.

1. Simulations in ideal sound-fields

Before examining the passive sonar scenarios detailed in Sec. IV A, a preliminary investigation of the expected advantages and limitations of the proposed method was performed. For the following tests, a complementary set of hydrophone array data is created with simulations of ideal plane waves and diffuse fields. The data contain the hydrophone signals of an array with the same geometry as the 96-channel array described in Sec. IV A, with the only difference being that the cylindrical baffle in this case has infinite length.

First, the HO-CroPaC post-filter was applied to the outputs of three beamformers: the DaS beamformer, a maximum-directivity CHB, and the MVDR beamformer. For this set of experiments, no added noise was included. The benefits of the proposed method are examined in terms

https://doi.org/10.1121/10.0005414

TABLE I. This table shows the values of directivity index (DI), angular resolution (RES), and maximum sidelobe level (MSL) averaged over frequency in the range of 5–9 kHz for different beamformers as well as the respective values obtained by post-filtering with the proposed HO-CroPaC method.

	DI (dB)	RES (deg)	MSL (dB)
DaS	18.6	4.2	−30
DaS-filtered	21.6	1.4	−46
CHB	19.5	2.8	−13.6
CHB-filtered	21.9	1.2	−37
MVDR	22.5	0.4	−42.8
MVDR-filtered	22.8	0.3	−54.8

of the following traditional array performance measures: DI, angular resolution (RES), and maximum sidelobe level (MSL). DI is calculated as the ratio (in dB) of the gain in the look-direction to the average gain over all directions; RES corresponds to the –3 dB width of the mainlobe, also known as half-power beamwidth; and MSL is the level difference between the peak of the highest sidelobe and the peak of the mainlobe. The results are provided in Table I. It can be observed that the post-filter can offer advantages in all three cases: for DaS and CHB, the average DI over the studied frequencies (5–9 kHz) is increased by 2–3 dB and the RES by about 1.5°–3°. The highest sidelobe is suppressed by 16 dB for DaS and by 23 dB for CHB. The advantages obtained for MVDR are less pronounced due to the exceptional performance of MVDR under ideal conditions (exact model, no added noise). However, as noted previously, the performance of MVDR degrades significantly at low SNR levels and, especially, in the presence of coherent interferers.

For the purpose of studying the performance of the method with background noise included, another set of simulated array data was created. In this case, a single source with fixed bearing at 0° was accompanied by a diffuse field, with a DDR from −50 to +20 dB in steps of 5 dB. The cylindrically isotropic diffuse field was simulated with 360 uniformly distributed incoherent-noise sources around the array. An example of the BTR plots obtained for DDR = –15 dB when the post-filter is applied to DaS is shown in Fig. 7. The level difference between the peak at 0° and the average output in other directions over this 10-s segment is denoted by D_1 for the baseline beamformer and D_2 for the filtered output. The advantage in terms of background noise attenuation obtained with post-filtering is therefore calculated as $D_a = D_2 - D_1$, which, for the shown case, is approximately 7.5 dB. By repeating this process for various DDR values and for the other two beamformers (CHB and MVDR), the curves shown in Fig. 8 are obtained. Figure 8(a) shows that post-filtering can always offer an advantage, so long as the peak is visible by the baseline beamformer. For the given array, the practical limit below which the peak is not visible by any beamformer ($D_1 < 1$ dB) is around DDR = –25 dB. From Fig. 8(b), it can be concluded that the proposed method can offer an advantage of approximately 17–20 dB when the ambient noise is moderately low (DDR = 20 dB). It is also interesting to note that all three methods have similar performance at very low DDR levels in terms of background noise attenuation. Note, however, that in reverberant environments with coherent reflections, the MVDR performance is expected to be more problematic, as demonstrated in Sec. IV C 2.

2. Passive sonar scenarios

In this section, the performance of the proposed method for the three passive sonar scenarios described in Sec. IV A is presented. The BTR plots are shown for the DaS beamformer and its filtered output with the HO-CroPaC filter. For each scenario, slices of the BTR plots are extracted at specific time instances of interest, where the results obtained by the MVDR beamformer and its filtered output are also plotted for comparison. The results obtained by the CHB beamformer are henceforth omitted, due to their high similarity with DaS, to improve graphical clarity.

The BTR plots for the first scenario, involving four moving sources, are shown in Fig. 9. The advantage can be directly inferred from the increased contrast in the second plot. A clearer depiction of the advantage gained by the proposed method is given in Fig. 10, which shows time-frame slices of the same BTR plots taken at the same time instance

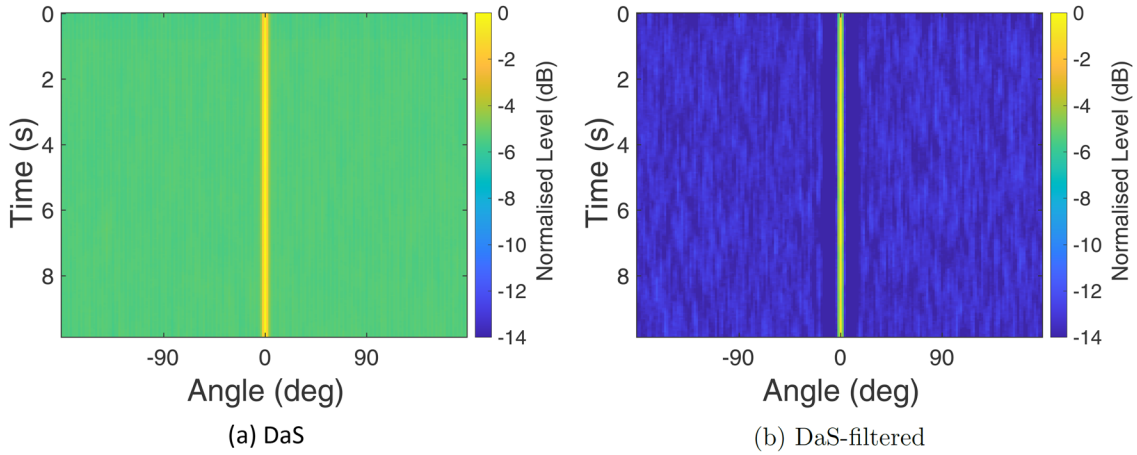


FIG. 7. (Color online) BTR plots of a source fixed at 0° in a diffuse field with DDR = –15 dB.

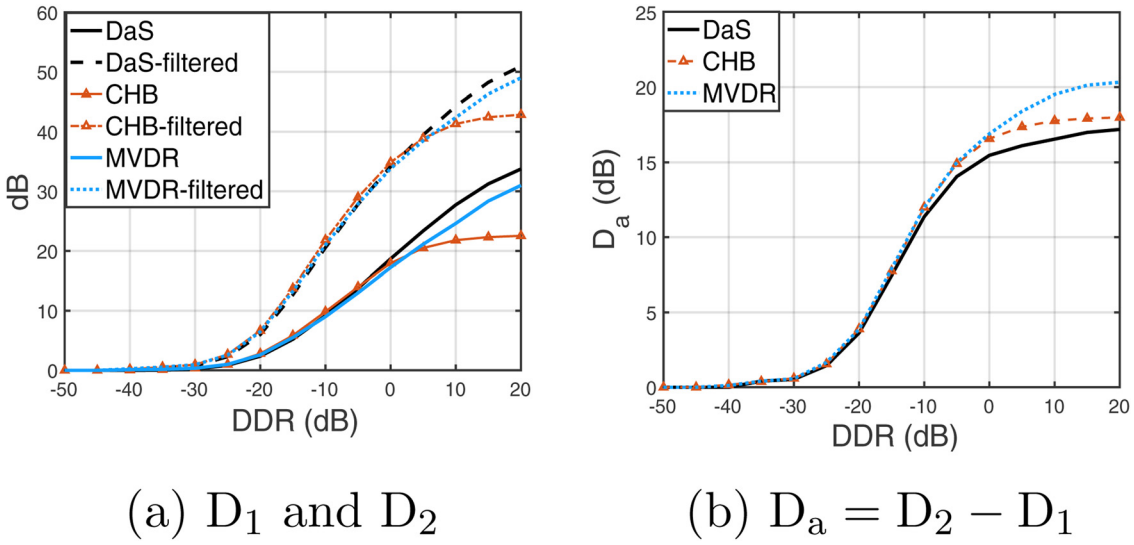


FIG. 8. (Color online) Average background noise attenuation in dB for the baseline beamformer (D_1: solid lines) and the filtered output (D_2: dashed and dotted lines) for different DDR values (a) and relative advantage $D_a = D_2 - D_1$ gained with the proposed method (b).

and plotted together for comparison. Here, it can be observed that post-filtering provides an additional attenuation of background noise of approximately 2.5 dB against the DaS. Moreover, the spatial resolution has increased due to the narrower mainlobe. The results obtained by MVDR and its filtered output are also depicted in the same plot for comparison. It can be seen that its performance is very close to that of DaS in this case. However, as validated in Sec. IV C 1, due to the very low SNR of this scenario, it is expected that the performance of MVDR should approach that of conventional beamforming (Zoltowski, 1988).

The results of the second scenario are shown in Figs. 11 and 12. The ambient noise in this case is significantly lower compared to the first scenario; however, the prominent acoustical reflection can be observed in the DaS output, along with some low-level sidelobe related aberrations. The sidelobes are attenuated by approximately 12–15 dB when using the HO-CroPaC filter. Note that since the reflection (sound source around -20°) is captured coherently by the HO-CroPaC directional patterns, it is not attenuated. However, both the reflection and the main source are still depicted with higher spatial resolution. The performance of MVDR is demonstrably worse than that of DaS for this scenario, despite the relatively high SNR level. However, since the lower-level signal is an acoustical reflection, whose signal is highly correlated with the main source signal, this was not unexpected. Other factors degrading MVDR's performance might include modeling errors, which do not affect DaS to the same extent (Cox, 1973). For example, as noted in Sec. IV B, modeling the baffle as an infinite cylinder is a deviation from the simulation; thus, this may have introduced errors in the steering vectors.

Note also the presence of dips around the two peaks in both filtered outputs. These non-linear artefacts are more dominant in cases of low background noise like the present scenario. They can be explained by the fact that when the beam patterns are steered in a direction adjacent to a source, the correlation of the signals is low, while the captured

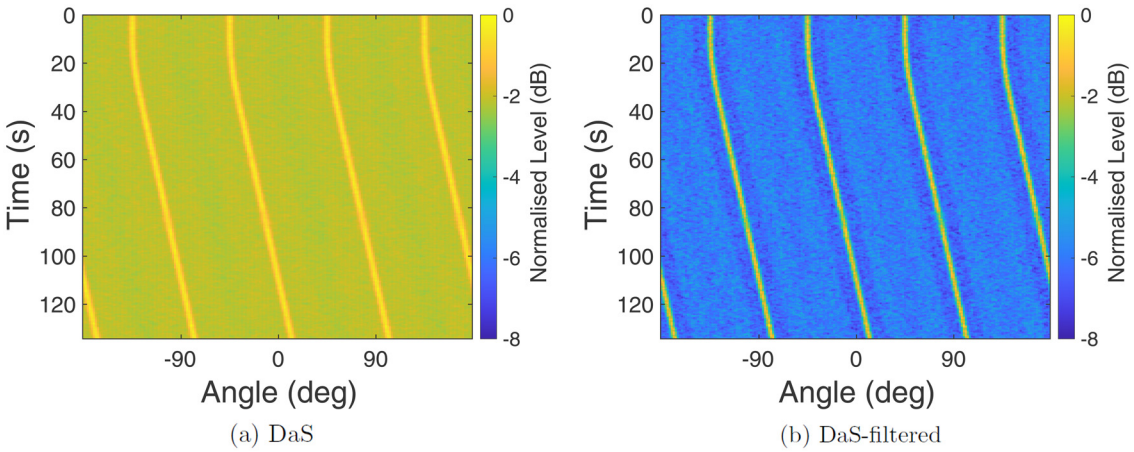


FIG. 9. (Color online) BTR plots of four moving sources in a diffuse field (scenario 1).

https://doi.org/10.1121/10.0005414

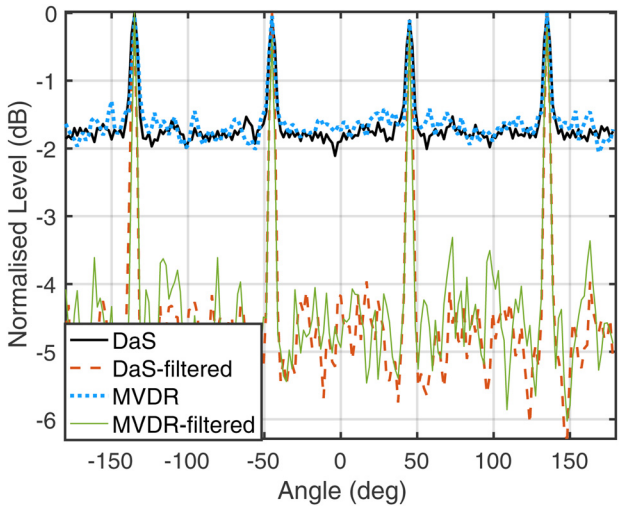


FIG. 10. (Color online) Comparison at a single time frame (scenario 1).

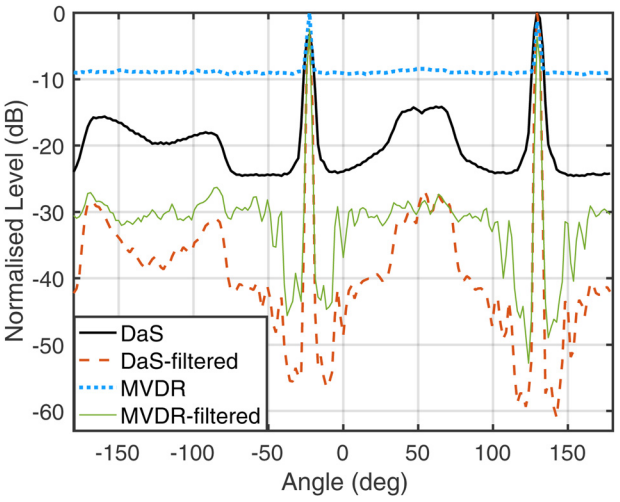


FIG. 12. (Color online) Comparison at a single time frame (scenario 2).

energy is high, as it includes much of the energy of the sound source. This results in very low G_r values in the vicinity of the peak. As a side note, it is possible to mitigate the effects of these non-linearities by setting a spectral floor on the computed G_r values, as suggested in Delikaris-Manias and Pulkki (2013) for the purpose of improving signal quality. However, this would also decrease the effectiveness of the post-filter in the other scanning directions and was therefore not deemed necessary in this case.

Finally, the BTR plots of the two closely spaced sources are shown in Fig. 13. Similar results to the previous scenario are observed regarding sidelobe attenuation and spatial resolution. Furthermore, the significance of the proposed method is clearly illustrated in Fig. 14(a), which shows the time frame where the DaS beamformer begins to fail at distinguishing between the two targets, whereas the filtered output succeeds in showing two distinct peaks. The non-linear artefacts around the peaks are also prominent in this scenario, due to the low ambient-noise energy. It is interesting to note that MVDR in this case is capable of better separating the two targets for the whole time segment. Figure 14(b) shows the time-frame plot obtained by MVDR and its filtered version at the same instant as the DaS. The dip between the peaks for DaS-filtered is approximately 2.5 dB, while for MVDR, the dip is approximately 7 dB. Nevertheless, the HO-CroPaC filter can offer an additional advantage by extending the dip to approximately 11.5 dB in this specific example. Moreover, the background noise level is attenuated by around 12 dB. It can be concluded that for a scenario of closely spaced sources with relatively high SNR, MVDR may be a preferable choice, while post-filtering may offer an additional advantage. However, due to the increase in the computational load (matrix inversions), along with the susceptibility to coherent reflections, DaS is likely to remain a popular choice in practice.

V. CONCLUSIONS

This study details a novel formulation of the CroPaC algorithm in the CHD. The main motivation for this reformulation is to improve the applicability of the method for very high CH orders ($N \gg 2$), where the original CroPaC formulation has historically suffered from the generation of significant sidelobes. The suppression of these lobes is

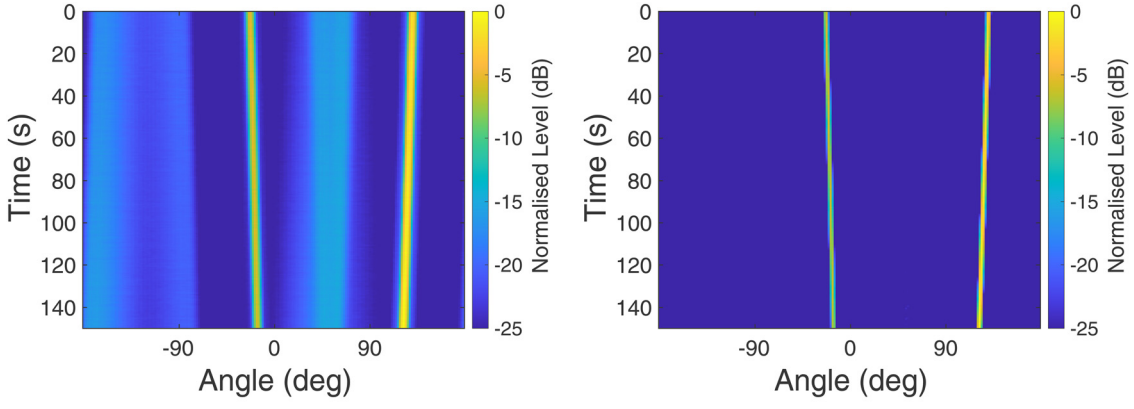
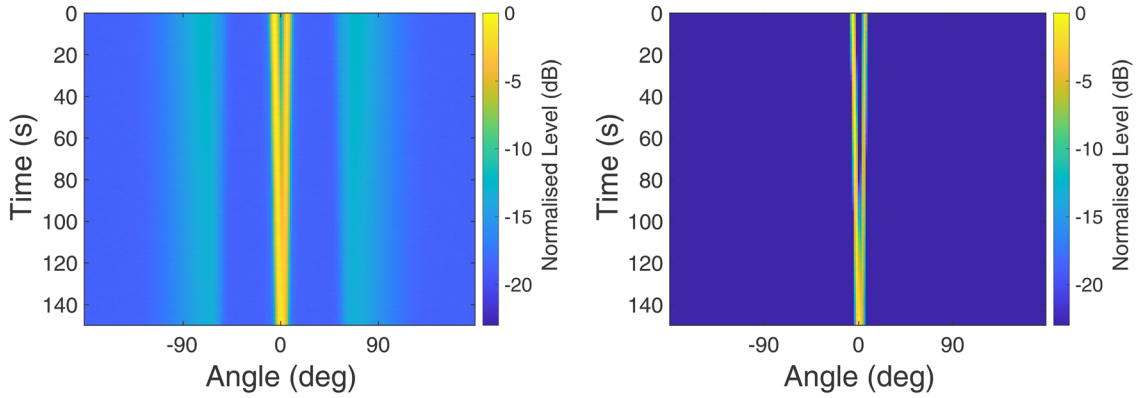
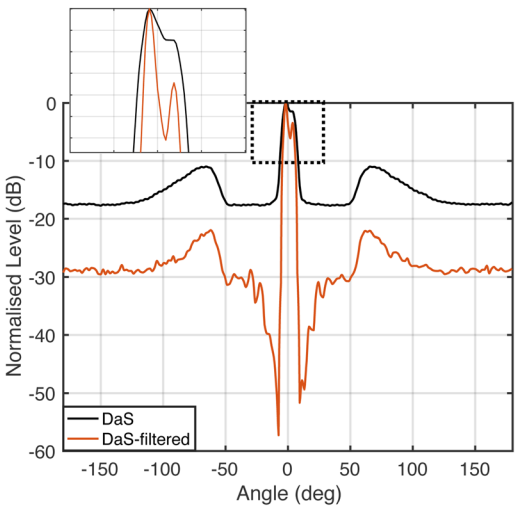


FIG. 11. (Color online) BTR plots of one moving source (scenario 2).

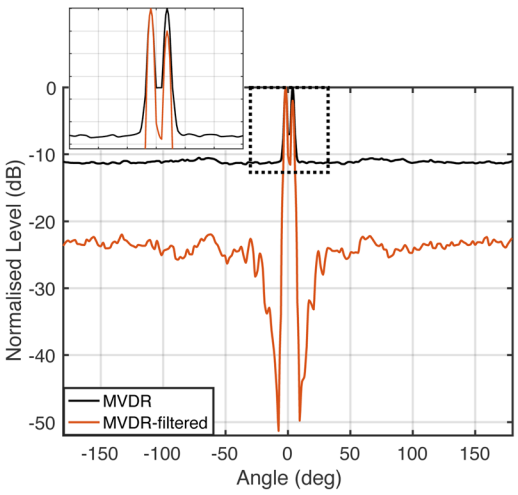


(a) DaS (b) DaS-filtered

FIG. 13. (Color online) BTR plots of two closely spaced sources (scenario 3).



(a) Delay-and-Sum as the baseline beamformer



(b) MVDR as the baseline beamformer

FIG. 14. (Color online) Comparison at a single time frame (scenario 3).

achieved by (a) designing the employed directional patterns to have same-phase sidelobes occurring in different directions and (b) spatially weighting them toward the look-direction with a cardioid beam-pattern. The computed time-frequency-dependent parameter may then be used to spatially filter signals captured by any beamformer, resulting in improved resolution as well as suppression of ambient noise and interfering sound sources outside the look-direction.

The proposed post-filter formulation is tested using simulated data of a circular 96-channel hydrophone array mounted onto a cylindrical baffle. The evaluation is performed both in ideal conditions and in realistic passive sonar scenarios employing highly detailed acoustical modeling. The results indicate an improved performance when applying the proposed post-filter to a conventional DaS beamformer, a CH beamformer, and an adaptive MVDR beamformer formulated in the CHD. In all studied cases, the proposed method can improve upon the resolution and background noise suppression capabilities of the baseline beamformers.

Overall, the results suggest that the reformulated CroPaC method can be effectively applied for underwater sound-field visualisation, a task that has high practical value for underwater surveillance, submarine navigation, and target localisation. Moreover, while this study was based on a formulation in the CHD, the findings suggest that it could be extended to the spherical harmonic domain for applications with spherical arrays.

ACKNOWLEDGMENTS

The authors would like to acknowledge SAAB AB for funding the project and SAAB Kockums AB for providing their expertise and the simulation data of the passive sonar scenarios.

Akaike, H. (**1974**). "A new look at the statistical model identification," IEEE Trans. Automat. Control **19**(6), 716–723.

Bernschütz, B., Pörschmann, C., Spors, S., Weinzierl, S., and der Verstärkung, B. (**2011**). "Soft-limiting of the modal amplitude amplification for spherical microphone arrays during the plane wave decomposition method," in *Proceedings of 37. German Annual Conference for Acoustics (DAGA 2011)*, March 21–24, Düsseldorf, Germany, pp. 661–662.

Bianchi, L., Verdi, M., Antonacci, F., Sarti, A., and Tubaro, S. (**2015**). "High resolution imaging of acoustic reflections with spherical microphone arrays," in *Proceedings of the 2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 18–21, New Paltz, NY.

Brinkmann, K., and Hurka, J. (**2009**). "Broadband passive sonar tracking," in *Informatik 2009—Im Focus das Leben*, edited by S. Fischer, E. Maehle, and R. Reischuk (German Informatics Society, Bonn, Germany).

Carter, G. (**1981**). "Time delay estimation for passive sonar signal processing," IEEE Trans. Acoust. Speech Signal Process. **29**(3), 463–470.

Chen, W., Wong, K. M., and Reilly, J. P. (**1991**). "Detection of the number of signals: A predicted eigen-threshold approach," IEEE Trans. Signal Process. **39**(5), 1088–1098.

Cox, H. (**1973**). "Resolving power and sensitivity to mismatch of optimum array processors," J. Acoust. Soc. Am. **54**(3), 771–785.

Delikaris-Manias, S., and Pulkki, V. (**2013**). "Cross pattern coherence algorithm for spatial filtering applications utilizing microphone arrays," IEEE Trans. Audio Speech Language Process. **21**(11), 2356–2367.

Delikaris-Manias, S., and Pulkki, V. (**2014**). "Cross spectral density based spatial filter employing maximum directivity beam patterns," in *Proceedings of IISA 2014*, the *5th International Conference on Information, Intelligence, Systems and Applications*, July 7–9, Chania, Greece, pp. 1–6.

Delikaris-Manias, S., Vilkamo, J., and Pulkki, V. (**2016**). "Signal-dependent spatial filtering based on weighted-orthogonal beamformers in the spherical harmonic domain," IEEE/ACM Trans. Audio Speech Language Process. **24**(9), 1511–1523.

Fahy, F., and Salmon, V. (**1990**). "Sound intensity," J. Acoust. Soc. Am. **88**, 2044–2045.

Ferguson, B. G., and Carevic, D. (**2010**). "Optimal spatial filtering of real data from submarine sonar arrays," in *Proceedings of OCEANS*'10 IEEE Sydney, May 24–27, Sydney, Australia, pp. 1–7.

Frost, O. L. (**1972**). "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**(8), 926–935.

Gao, F., and Gershman, A. B. (**2005**). "A generalized esprit approach to direction-of-arrival estimation," IEEE Signal Process. Lett. **12**(3), 254–257.

Habets, E. A., Benesty, J., Gannot, S., Naylor, P. A., and Cohen, I. (**2009**). "On the application of the LCMV beamformer to speech enhancement," in *Proceedings of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 18–21, New Paltz, NY, pp. 141–144.

Han, K., and Nehorai, A. (**2013**). "Improved source number detection and direction estimation with nested arrays and ULAs using jackknifing," IEEE Trans. Signal Process. **61**(23), 6118–6128.

Ivanic, J., and Ruedenberg, K. (**1998**). "Rotation matrices for real spherical harmonics. Direct determination by recursion," J. Phys. Chem. A **102**(45), 9099–9100.

Jin, C. T., Epain, N., and Parthy, A. (**2014**). "Design, optimization and evaluation of a dual-radius spherical microphone array," IEEE/ACM Trans. Audio Speech Language Process. **22**(1), 193–204.

Jo, B., Zotter, F., and Choi, J.-W. (**2020**). "Extended vector-based eb-esprit method," IEEE/ACM Trans. Audio Speech Language Process. **28**, 1692–1705.

Liu, Y. (**2012**). "Phase modes circular array superdirective beamforming," in *Proceedings of 2012 OCEANS*, October 14–19, Hampton Roads, VA, pp. 1–6.

McCormack, L., Delikaris-Manias, S., Politis, A., Pavlidi, D., Farina, A., Pinardi, D., and Pulkki, V. (**2019**). "Applications of spatially localized active-intensity vectors for sound-field visualization," J. Audio Eng. Soc. **67**(11), 840–854.

McCormack, L., Delikaris-Manias, S., and Pulkki, V. (**2017**). "Parametric acoustic camera for real-time sound capture, analysis and tracking," in *Proceedings of the 20th International Conference on Digital Audio Effects (DAFx-17)*, September 5–9, Edinburgh, UK, pp. 412–419.

McCowan, I. A., and Bourlard, H. (**2003**). "Microphone array post-filter based on noise field coherence," IEEE Trans. Speech Audio Process. **11**(6), 709–716.

McIntyre, G., Loadman, C., Bousquet, J.-F., and Blouin, S. (**2015**). "Low power beamforming for underwater acoustic sensing using a 5-element circular hydrophone array," in *Proceedings of 2012 OCEANS 2015–Genova*, May 18–21, Genova, Italy, pp. 1–8.

Meyer, J. (**2001**). "Beamforming for a circular microphone array mounted on spherically shaped objects," J. Acoust. Soc. Am. **109**(1), 185–193.

Moore, A. H., Evers, C., and Naylor, P. A. (**2017**). "Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors," IEEE/ACM Trans. Audio Speech Language Process. **25**(1), 178–192.

Moreau, S., Daniel, J., and Bertet, S. (**2006**). "3D sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *Proceedings of the 120th Convention of the AES*, May 20–23, Paris, France, pp. 20–23.

Parthy, A., Epain, N., van Schaik, A., and Jin, C. T. (**2011**). "Comparison of the measured and theoretical performance of a broadband circular microphone array," J. Acoust. Soc. Am. **130**(6), 3827–3837.

Peled, Y., and Rafaely, B. (**2011**). "Linearly constrained minimum variance method for spherical microphone arrays in a coherent environment," in *Proceedings of the 2011 Joint Workshop on Hands-Free Speech Communication and Microphone Arrays*, May 30–June 1, Edinburgh, UK, pp. 86–91.

Pirkl, R. J., and Aughenbaugh, J. M. (**2015**). "Bayesian passive sonar tracking with conventional beamformer-level data," in *Proceedings of the 2015 18th International Conference on Information Fusion (Fusion)*, July 6–9, Washington, DC, pp. 621–628.

Politis, A., and Gamper, H. (**2017**). "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in *2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, October 15–18, New Paltz, NY, pp. 224–228.

Pulkki, V., Delikaris-Manias, S., and Politis, A. (**2018**). *Parametric Time-Frequency Domain Spatial Audio* (Wiley, New York).

Quazi, A. (**1981**). "An overview on the time delay estimate in active and passive systems for target localization," IEEE Trans. Acoust. Speech Signal Process. **29**(3), 527–533.

Rafaely, B. (**2005**). "Phase-mode versus delay-and-sum spherical microphone array processing," IEEE Signal Process. Lett. **12**(10), 713–716.

Reddy, V., Paulraj, A., and Kailath, T. (**1987**). "Performance analysis of the optimum beamformer in the presence of correlated sources and its behavior under spatial smoothing," IEEE Trans. Acoust. Speech Signal Process. **35**(7), 927–936.

Roy, R., and Kailath, T. (**1989**). "Esprit-estimation of signal parameters via rotational invariance techniques," IEEE Trans. Acoust. Speech Signal Process. **37**(7), 984–995.

Schmidt, R. (**1986**). "Multiple emitter location and signal parameter estimation," IEEE Trans. Antennas Propag. **34**(3), 276–280.

Simmer, K. U., Bitzer, J., and Marro, C. (**2001**). "Post-filtering techniques," in *Microphone Arrays* (Springer, New York), pp. 39–60.

Tervo, S., and Politis, A. (**2015**). "Direction of arrival estimation of reflections from room impulse responses using a spherical microphone array," IEEE/ACM Trans. Audio Speech Language Process. **23**(10), 1539–1551.

Teutsch, H. (**2007**). *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*, 348 (Springer, New York).

Teutsch, H., and Kellermann, W. (**2006**). "Acoustic source detection and localization based on wavefield decomposition using circular microphone arrays," J. Acoust. Soc. Am. **120**(5), 2724–2736.

Tiana-Roig, E., Jacobsen, F., and Grande, E. F. (**2010**). "Beamforming with a circular microphone array for localization of environmental noise sources," J. Acoust. Soc. Am. **128**(6), 3535–3542.

Torres, A. M., Cobos, M., Pueo, B., and Lopez, J. J. (**2012**). "Robust acoustic source localization based on modal beamforming and time–frequency processing using circular microphone arrays," J. Acoust. Soc. Am. **132**(3), 1511–1520.

Van Trees, H. L. (**2004**). *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory* (Wiley, New York), pp. 17–322.

Van Veen, B. D., and Buckley, K. M. (**1988**). "Beamforming: A versatile approach to spatial filtering," IEEE ASSP Mag. **5**(2), 4–24.

Williams, E. G. (**1999**). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography* (Academic, New York).

Yan, S. (**2015**). "Optimal design of modal beamformers for circular arrays," J. Acoust. Soc. Am. **138**(4), 2140–2151.

Yan, S. (**2020**). "Robust time-domain broadband modal beamforming for circular arrays," IEEE Trans. Aerosp. Electron. Syst. **56**(3), 1783–1794.
Zoltowski, M. D. (**1988**). "On the performance analysis of the MVDR beamformer in the presence of correlated interference," IEEE Trans. Acoust. Speech Signal Process. **36**(6), 945–947.
Zotter, F., and Frank, M. (**2019**). *Ambisonics* (Springer, New York), pp. 175–178.
Zou, N., and Nehorai, A. (**2009**). "Circular acoustic vector-sensor array for mode beamforming," IEEE Trans. Signal Process. **57**(8), 3041–3052.