Audio Engineering Society

Convention Paper

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

Real-time underwater spatial audio: a feasibility study

Symeon Delikaris-Manias, Leo McCormack, Ilkka Huhtakallio, and Ville Pulkki

Aalto Acoustics Lab, Aalto University, Espoo, Finland

Correspondence should be addressed to Symeon Delikaris-Manias (symeon.delikarismanias@gmail.com)

ABSTRACT

In recent years, spatial audio utilising compact microphone arrays has seen many advancements due to emerging virtual reality hardware and computational advances. These advances can be observed in three main areas of spatial audio, namely: spatial filtering, direction of arrival estimation and sound reproduction over loudspeakers or headphones. The advantage of compact microphone arrays is their portability, which permits their use in everyday consumer applications. However, an area that has received minimal attention is the field of underwater spatial audio, using compact hydrophone arrays. Although the principles are largely the same, microphone array technologies have rarely been applied to underwater acoustic arrays. In this feasibility study we present a purpose built compact hydrophone array, which can be transported by a single diver. This study demonstrates a real-time underwater acoustic camera for underwater sound-field visualisation and a parametric binaural rendering engine for auralisation.

1 Introduction

This paper presents a feasibility study regarding the use of compact hydrophone arrays for underwater spatial audio applications. An open-body tetrahedral array was built, consisting of four hydrophones placed at the vertices of a tetrahedral frame. The hydrophone array signals were then spatially encoded into spherical harmonic signals [1] (also known as B-Format or Ambisonic signals) using theoretical filters in real-time. Such an approach has only sparsely been observed in the literature for underwater applications [2, 3, 4, 5]. Whereas, this approach has been utilised by many commercial microphone arrays, which can attain high spatial coherence in the resulting spherical harmonic signals [6]. They are often utilised for Ambisonics reproduction, [7], or in non-linear parametric spatial audio reproduction systems [8].

In this study, the spherical harmonic signals derived from the hydrophone signals were subsequently utilised by various sound-field visualisation algorithms and a parametrically enhanced Ambisonic binaural reproduction method. The sound-field visualisation is represented as an activity map, similar to the result of a scanning beamformer; where the relative energy of beamformers for many directions is depicted using a colour gradient. Two algorithms were programmed into an audio plug-in to visualise the sound field: the first utilises the Cross-Pattern Coherence (CroPaC) spatial filter [9], while the second is based on a variant of the multiple signal classification (MUSIC) algorithm [10]. A number of experiments were then performed

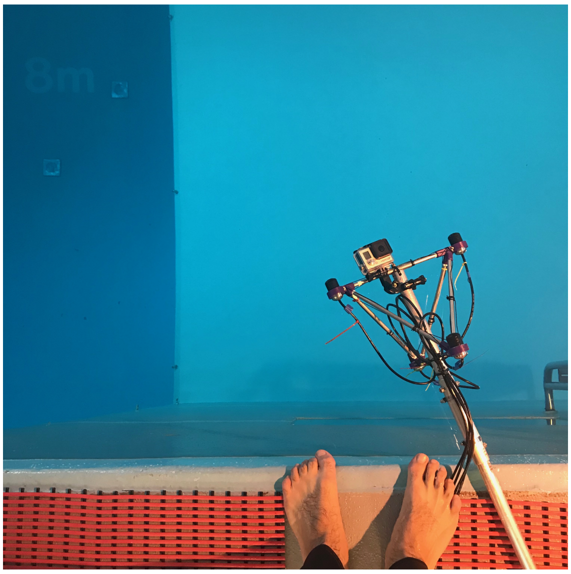


Fig. 1: The hydrophone array consisting of four hydrophones in a tetrahedral arrangement and a waterproof video camera to provide the corresponding video stream.

in a diving pool with the tetrahedral array and an underwater loudspeaker; including a realistic scenario of tracking the position of a moving diver in real-time. Visualisation of the sound-field inside the pool, along with the corresponding video footage, is also presented from the perspective of the array; where the visualisation software is based on a parametric acoustic camera detailed in [11]. Auralisation is achieved by utilising a first-order formulation of the most recent binaural Directional Audio Coding (DirAC) method [12], which adaptively mixes the Ambisonic decoded audio with the parametrically enhanced DirAC stream [13]. This approach attempts to retain the high single-channel sound quality of the linear decoded audio, while simultaneously improving the perceived spatial accuracy.

After conducting experiments in a diving pool, this work suggests that this simple tetrahedral hydrophone array can be utilised in real-time to identify the direction of divers with relatively high accuracy; both visually and aurally. The array, along with the feet of the diver, are depicted in Fig. 1.

This paper is organised as follows: Section 2 details the construction of the tetrahedral hydrophone array; Section 3 provides background regarding how the hydrophone array signals were spatially encoded into first-order spherical harmonic signals; Section 4 describes how these spherical harmonic signals were utilised to visualise the sound-field; Section 5 details a the first-order DirAC formulation, utilised for headphone reproduction of the sound-field; Section 6 describes the experiment, which investigated the feasibility of utilising these aforementioned techniques for underwater use; and Section 7 concludes the paper.

2 Hydrophone array construction

The hydrophone array was designed and built based on four Aquarian-Audio H2a-XLR hydrophones. The shape of the hydrophone is a spherocylindrical capsule with a diameter of 24.9ṁm and length of 45.5ṁm. The hydrophone capsule includes the transducer and buffer pre-amplifier, which is powered by a standard 48V̇ Phantom power feed via a balanced signal cable with an XLR connector at the dry end. Each of the four hydrophones in the array had its own independent cable and P48 feed. The manufacturer stated sensitivity for the hydrophones is within +/-4ḋB from 20Ḣz to 4k̇Hz and the directional behaviour is assumed to be omnidirectional within the operating band for this application. The linearity or directional characteristics of the hydrophones were not verified by the authors.

The tetrahedral array was built from custom designed and 3D-printed corner clamps and stainless steel rods. The corner clamps hold the hydrophones tightly in place and the rods connected the corner clamps together. ABS was used as the 3D-printing material. The axes of the hydrophones were parallel and the radius of the sphere connecting the centres of the hydrophones was approximately 173ṁm. A computer aided design (CAD) image of the hydrophone array is depicted in Fig. 2.

The array design differs from one other hydrophone array proposed recently in [4], which consisted of four sensors orientated in a line array. Such a design is suitable for beamforming and direction-of-arrival (DOA) estimation on a 2-D plane and the authors presented practical means of compensating for mismatches between the sensors. Whereas for the proposed tetrahedral array, its spatial characteristics were assumed to behave in the same manner as described by their theoretical modal counterparts [14, 15]; and the frequency

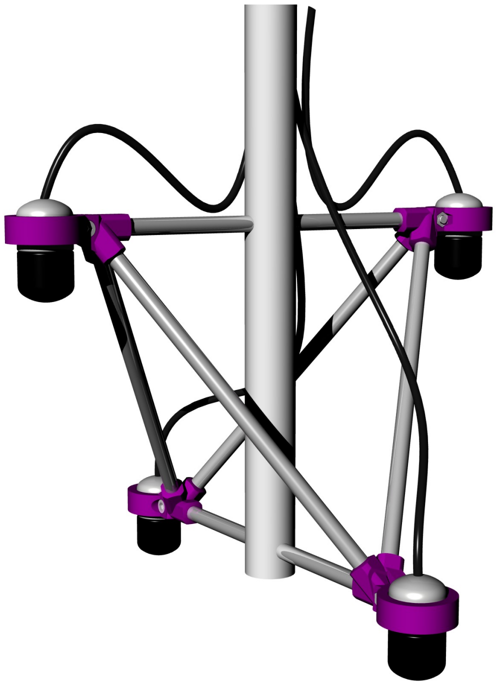


Fig. 2: A CAD image of the hydrophone array.

response was assumed to be the same for each sensor. However, as demonstrated later in the paper, the performance of the array was sufficient for tracking the underwater movements of a diver in 3-D and for plausible auralisation via headphones.

3 Spatial encoding

The hydrophone array has $Q = 4$ sensors at $\Omega_q = (\theta, \phi, r)$ locations; where $\theta \in [-\pi/2, \pi/2]$ denotes the elevation angles, $\phi \in [-\pi, \pi]$ the azimuthal angles and r the radius of the tetrahedron. In order to make the hydrophone array signals $\mathbf{x} \in \mathbb{C}^{Q\times 1}$ suitable for use with the algorithms described later in the paper, they must first be decomposed into first-order spherical harmonic signals; where the accuracy of this decomposition is largely dependent on the sensor distribution and the radius of the array [1]. This spatial encoding is performed in the time-frequency domain, with time and frequency indices denoted with (t, f).

The spherical harmonic signals can be estimated as

$$\mathbf{s}(t,f) = \mathbf{E}(f)\mathbf{x}(t,f), \tag{1}$$

where

$$\mathbf{s}(t,f) = [s_{00}(t,f), s_{1-1}(t,f), ..., s_{11}(t,f)]^T \in \mathbb{C}^{4\times 1}, \tag{2}$$

are the spherical harmonic signals [1] and $\mathbf{E}(f) \in \mathbb{C}^{4\times 4}$ is a frequency-dependent spatial encoding matrix. Assuming a uniform distribution of sensors, it can be formulated as

$$\mathbf{E}(f) = \frac{1}{Q}\mathbf{W}(f)\mathbf{Y}, \tag{3}$$

where $\mathbf{Y} \in \mathbb{R}^{4\times 4}$ is a matrix containing the spherical harmonics weights for each sensor direction and $\mathbf{W}(f) \in \mathbb{C}^{4\times 4}$ is a diagonal equalisation matrix.

The frequency-independent spherical harmonics matrix is given as

$$\mathbf{Y} = \begin{bmatrix} Y_{00}(\Omega_1) & Y_{1-1}(\Omega_1) & Y_{10}(\Omega_1) & Y_{11}(\Omega_1) \\ Y_{00}(\Omega_2) & Y_{1-1}(\Omega_2) & Y_{10}(\Omega_2) & Y_{11}(\Omega_2) \\ Y_{00}(\Omega_3) & Y_{1-1}(\Omega_3) & Y_{10}(\Omega_3) & Y_{11}(\Omega_3) \\ Y_{00}(\Omega_4) & Y_{1-1}(\Omega_4) & Y_{10}(\Omega_4) & Y_{11}(\Omega_4) \end{bmatrix}, \tag{4}$$

where Y_{lm} are the real-valued spherical harmonics of order $l \geq 0$, degree $m \in [-l, l]$ and direction Ω.

The diagonal equalisation matrix is defined as

$$\mathbf{W}(f) = \text{diag}\{[w_0(f), w_1(f), w_1(f), w_1(f)]\}, \tag{5}$$

where $w_l(f)$ are the order-dependent equalisation weights, which are the regularised inverse of the theoretical modal coefficients b_l using, for example, the Tikhonov method [6]

$$w_l(f) = \frac{1}{b_l(f)} \frac{|b_l(f)|^2}{|b_l(f)|^2 + \lambda^2}, \tag{6}$$

where λ is a regularisation parameter that allows a compromise between noise amplification and the accuracy of the transform. For an open spherical array with omnidirectional sensors, the modal coefficients are calculated as [15]

$$b_l(f) = 4\pi i^l j_l(kr), \tag{7}$$

where $i^2 = -1$, $j_l(.)$ is the spherical Bessel function of the first kind, $k = 2\pi c/f$ is the wave number and c is the speed of sound. A detailed description of the spatial encoding for microphone arrays can be found in [16].

[1] The spherical harmonic conventions utilised in this work are orthonormalised (N3D) real spherical harmonics with the Ambisonic Channel Number (ACN) indexing.

4 Visualisation

The sound-field is visualised by calculating the MUSIC pseudo-spectrum and a CroPaC parameter for a dense spherical grid. Both methods operate in the time-frequency domain and are capable of providing a more robust performance than traditional scanning beamformers [11]. A description of the principles for both methods is given in this section.

4.1 Pseudo-spectrum based sound field visualisation

The first step is to calculate the covariance matrix of the spherical harmonic signals, $\mathbf{C}_{lm}(f) = \mathrm{E}[\mathbf{s}\mathbf{s}^H] \in \mathbb{C}^{4\times 4}$, per frequency. A singular value decomposition (SVD) is then utilised to decompose the covariance matrix into the signal and noise subspaces

$$\mathbf{C}_{lm} = \mathbf{U}\mathbf{V}\mathbf{U}^H = [\mathbf{U}_s\mathbf{U}_n]\begin{bmatrix}\mathbf{V}_s & \mathbf{0}\\ \mathbf{0} & \mathbf{V}_n\end{bmatrix}\begin{bmatrix}\mathbf{U}_s\\ \mathbf{U}_n\end{bmatrix}, \quad (8)$$

where $\mathbf{U}_s \in \mathbb{C}^{1\times 1}$ and $\mathbf{U}_n \in \mathbb{C}^{3\times 3}$ are the signal subspace and the noise subspace, respectively; and $\mathbf{V}$ denotes the singular values. The time-frequency bins that are assumed to be from the direct path of a sound source to a receiver are obtained via a direct-path dominance test. They are selected by determining whether the first singular value of matrix $\mathbf{V}$ is significantly larger than the second singular value. This selection is subject to a threshold. The pseudo-spectrum is then obtained as

$$S_{\mathrm{MAP}}(\Omega_j) = \frac{1}{\mathbf{y}(\Omega_j)\Big(\mathbf{I} - \mathbf{U}_s\mathbf{U}_s{}^H\Big)\mathbf{y}^H(\Omega_j)}, \quad (9)$$

where $\mathbf{y}$ are the spherical harmonic weights for direction Ω_j, S_{MAP} is corresponding pseudo-spectrum value, and $\mathbf{I}$ is an identity matrix.

4.2 Parameter-based sound-field visualisation

The parametric sound-field visualisation approach utilised in this work is based on the CroPaC parameter, which essentially estimates a probability-like parameter of an active sound source being received from a specific direction [9]. In this implementation, CroPaC is based on calculating the cross-spectrum between two spherical harmonic signals that have the same look direction Ω_j

$$G(\Omega_j,t,f) = \hat{\lambda}\,\frac{\Re[s_{00}(\Omega_j,t,f)^* s_{11}(\Omega_j,t,f)]}{|s_{00}(\Omega_j,t,f)|^2 + |s_{11}(\Omega_j,t,f)|^2}, \quad (10)$$

where $\Re$ is the real operator, * denotes the complex conjugate and $\hat{\lambda} = \frac{5}{2}$ [11].

The CroPaC parameter is estimated for a spherical grid of look directions $\Omega = (\Omega_1, \Omega_2, \ldots, \Omega_J)$ by rotating the sound-field using an appropriate spherical harmonic rotation matrix [17]. The resulting activity-map is then given by

$$G_{\mathrm{MAP}}(\Omega_j,t) = \max\Big[0, \frac{1}{F}\sum_{f=1}^{F} G(\Omega_j,t,f)\Big], \quad (11)$$

where F is the total number of frequency bands. The half-wave rectification process ensures that only sounds arriving from the look direction are analysed; for more information on the method, the reader is directed to [9, 18].

5 Auralisation

For the binaural reproduction of the first-order spherical harmonic signals, a simplified version of the DirAC formulation detailed in [12], was utilised. The first step is to transform the signals into the time-frequency domain, via a short-time Fourier transform (STFT) (or alternatively, a perfect reconstruction filterbank). This is then followed by preliminary ambisonic decoding to headphones utilising a matrix of filters $\mathbf{D}(f) \in \mathbb{C}^{2\times 4}$, which are based on a set of Head-Related Transfer Functions (HRTFs)

$$\mathbf{y}_{\mathrm{lin}}(t,f) = \mathbf{D}(f)\mathbf{s}(t,f). \quad (12)$$

This ambisonic decoding matrix can be designed using a virtual loudspeaker approach [19, 20], or by expressing the HRTFs in the spherical harmonic domain [21]; for more details the reader is directed to [22].

Optionally, in order to accommodate head-tracking, the first-order spherical harmonic signals can be rotated via an appropriate spherical harmonic domain rotation matrix $\mathbf{M}_{\mathrm{rot}} \in \mathbb{R}^{4\times 4}$ [17]

$$\mathbf{y}_{\mathrm{lin}}^{(\mathrm{rot})}(t,f) = \mathbf{D}(f)\mathbf{M}_{\mathrm{rot}}(\alpha,\beta,\gamma)\mathbf{s}(t,f), \quad (13)$$

where α, β, γ are the rotation angles received from the head-tracker.

5.1 DirAC analysis

The next step in the DirAC method is to extract perceptually meaningful parameters from this spherical harmonic representation of the sound-field; namely the DOA, energy density and diffuseness. The omnidirectional and dipole signals, described by the zeroth and first-order spherical harmonic signals, respectively, relate directly to the pressure $p(t,f)$ and three-dimensional particle-velocity $\mathbf{u}(t,f) \in \mathbb{C}^{3\times 1}$ [8]. Therefore, the active intensity vector $\mathbf{i}_{\mathrm{a}}(t,f) \in \mathbb{C}^{3\times 1}$, which expresses the flow of acoustical energy, can be utilised to ascertain a DoA estimate per time and frequency index, and is estimated as

$$\mathbf{i}_{\mathrm{a}}(t,f) = \Re[p(t,f)^{*}\mathbf{u}(t,f)], \tag{14}$$

and the DoA, Φ is defined as

$$\Phi(t,f) = \angle\, \mathrm{E}\left[\frac{\mathbf{i}_{\mathrm{a}}(t,f)}{||\mathbf{i}_{\mathrm{a}}(t,f)||}\right], \tag{15}$$

where $\angle$ is the angle of a three-dimensional vector and E is the expectation operator.

The diffuseness parameter ψ can then be obtained as

$$\psi(t,f) = 1 - \frac{|\mathrm{E}[\mathbf{i}_{\mathrm{a}}(t,f)]|}{c\,\mathrm{E}[e(t,f)]}, \tag{16}$$

where e is the energy density, calculated as

$$e(t,f) = \frac{\rho_0}{2}|\mathbf{u}(t,f)|^2 + \frac{|p(t,f)|^2}{2\rho_0\,c^2} \tag{17}$$

where ρ_0 is the density of the medium.

These spatial parameters are then utilised to define a target covariance matrix $\mathbf{C}_{\mathrm{target}}(f) \in \mathbb{C}^{2\times 2}$, which is accumulated over a sufficiently long temporal window T, as to provide meaningful signal statistics

$$\mathbf{C}_{\mathrm{target}}(f) = \mathbf{C}_{\mathrm{dir}}(f) + \mathbf{C}_{\mathrm{diff}}(f) \tag{18}$$

where $\mathbf{C}_{\mathrm{dir}}(f) \in \mathbb{C}^{2\times 2}$ and $\mathbf{C}_{\mathrm{diff}}(f) \in \mathbb{C}^{2\times 2}$, refer to the direct and diffuse contributions to the target covariance matrix respectively, and are defined as [12]

$$\mathbf{C}_{\mathrm{dir}}(f) = \sum_{t}^{T} e(t,f)[1-\psi(t,f)]\,\mathbf{h}(\Phi,f)\mathbf{h}^{H}(\Phi,f), \tag{19}$$

$$\mathbf{C}_{\mathrm{diff}}(f) = \sum_{t}^{T} e(t,f)\mathbf{U}(f)\psi(t,f), \tag{20}$$

where $\mathbf{h}(\Phi,f) \in \mathbb{C}^{2\times 1}$ refers to the HRTFs that correspond to the analysed DoAs per time-frequency tile, and $\mathbf{U} \in \mathbb{R}^{2\times 2}$ denotes a matrix that integrates a binaural coherence curve, which defines how the diffuse energy should be distributed for headphone playback; for more details on calculating this matrix, the reader is directed to [23, 24].

5.2 DirAC synthesis

Unlike in the original (2006) DirAC formulation [25], the DirAC synthesis is now compromised of an *optimal mixing* between the linear ambisonic decoding and the parametric DirAC decoding [26]. Essentially, this approach attempts to derive mixing matrices which, when applied to the ambisonic decoded signals, brings the covariance matrix of the resulting signals closer to the target covariance matrix. Since the binaural cues utilised for human spatial sound perception relate directly to the inter-channel level, phase and coherence differences captured by the narrowband covariance matrix of the binaural signals, the spatial accuracy of the reproduced signals should be enhanced; while simultaneously preserving much of the high single-channel sound quality provided by the Ambisonics method.

This optimal mixing problem is defined as

$$\mathbf{y}_{\mathrm{bin}} = \mathbf{A}\mathbf{y}_{\mathrm{lin}} + \mathbf{B}\mathscr{D}[\mathbf{y}_{\mathrm{lin}}], \tag{21}$$

where $\mathscr{D}[.]$ refers to a decorrelation operation on the enclosed signals, and the mixing matrices $\mathbf{A} \in \mathbb{C}^{2\times 2}, \mathbf{B} \in \mathbb{C}^{2\times 2}$ are the solution to

$$\mathbf{A}\mathbf{C}_{\mathrm{lin}}\mathbf{A}^{H} + \mathbf{B}\tilde{\mathbf{C}}_{\mathrm{lin}}\mathbf{B}^{H} = \mathbf{C}_{\mathrm{target}}, \tag{22}$$

where $\mathbf{C}_{\mathrm{lin}} = \mathrm{E}[\mathbf{y}_{\mathrm{lin}}\mathbf{y}_{\mathrm{lin}}^{H}] \in \mathbb{C}^{2\times 2}$ is the covariance matrix of the ambisonic decoded signals and $\tilde{\mathbf{C}}_{\mathrm{lin}} = \mathrm{diag}\big(\mathrm{E}[\mathscr{D}[\mathbf{y}_{\mathrm{lin}}]\,\mathscr{D}[\mathbf{y}_{\mathrm{lin}}]^{H}]\big) \in \mathbb{C}^{2\times 2}$ is a diagonal matrix, defined as the diagonal elements of the covariance matrix of the decorrelated decoded signals.

Essentially, the solution first tries to impose the energies and coherences defined by $\mathbf{C}_{\mathrm{target}}$, via a linear mapping using mixing matrix $\mathbf{A}$; therefore, minimising decorrelation artefacts, such as temporal smearing of transients, as much as possible. Mixing matrix $\mathbf{B}$ is then utilised to fulfil the remaining target dependencies. For a more detailed explanation and for the derivation of these mixing matrices, the reader is referred to [13, 27].

6 Experiment

The experiments were performed in a diving pool located in Kirkkonummi, Finland. The pool was L-shaped with horizontal dimensions of the longer leg 5 x 12m and the shorter leg 5 x 8m. The depth of the pool was 5 meters, with the exception of the corner where the legs met; where the depth was 8 meters.

The depth of the hydrophone array was fixed by attaching it to the end a metal bar, which was made out of a three pieces of 1.5m long aluminium tubes. The length of the bar could be controlled by removing or adding additional pieces. The bar was fed through a large fishing buoy and the point where it was fixed, determined the depth of the array; thus allowing the vertical placement of the array to be controlled. The horizontal positioning and orientation was controlled by attaching the assembly to the end of a 8m long fishing rod, which was operated from the side of the pool. The rod worked also as a bridge for the four signal cables. On the dry end, the signal acquisition was handled using a RME Micstasy preamp / AD-converter and RME Fireface 400 interface. The Micstasy also provided the P48 phantom power for the hydrophones.

The hydrophone array signals were encoded into spherical harmonic signals using an audio plug-in detailed in [28]; which is part of the upcoming Spatial Audio Real-Time Applications (SPARTA) plugins[2], where a maximum of 15dB amplification using the Tikhonov regularisation approach was utilised. The MUSIC and CroPaC activity map visualisations were then obtained using the first-order spherical harmonic signals via the acoustic camera audio plug-in developed in [11], and the auralisations were performed by an audio plug-in implementation of the latest DirAC formulation [12].

After informal testing it was found that an observer was able to clearly track the movements of the diver from the perspective of the array, both visually and aurally[3]. The diver from the perspective of the array and the corresponding activity map are shown in Fig. 3 and Fig. 4, respectively.

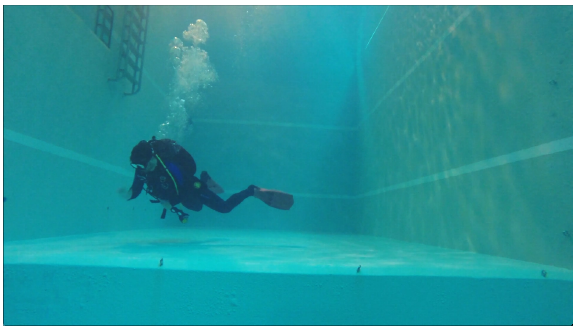
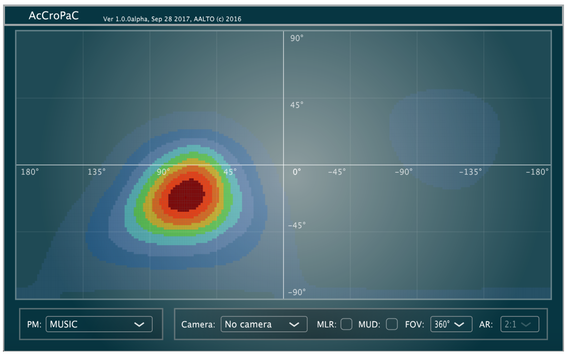


Fig. 3: The diver from the perspective of the hydrophone array.



AcCroPaC Ver 1.0.0alpha, Sep 28 2017, AALTO (c) 2016

PM: MUSIC

Camera: No camera MLR: □ MUD: □ FOV: 360° AR: 2:1

Fig. 4: The corresponding activity map using the acoustic camera.

Note however, that due to the low spatial-resolution of the first-order signals, the sub-space MUSIC approach provided clearer activity maps than the CroPaC approach. However, as shown in [9, 11], the CroPaC method has been shown be more robust in more reverberant and noisy environments with higher spatial resolution input signals; therefore, it is important to have the ability to switch between the two methods to better suit the environment and input resolution.

7 Summary

This paper has detailed a tetrahedral hydrophone array for underwater spatial audio applications. Some of the state-of-the-art algorithms utilised today for compact microphone arrays were investigated. This feasibility study demonstrates the real-time encoding of hydrophone array signals in the intermediate format of spherical harmonics signals and the visualization and auralisation of an underwater environment. The underwater sound-field was visualised as an activity map using two types of signal-dependent techniques: a

[2] `http://research.spa.aalto.fi/projects/sparta_vsts/`

[3] A video footage of the experiment can be found here: `https://youtu.be/3WARepl3lEg`

subspace-based (MUSIC) and a cross-spectrum-based (CroPaC). The auralisation is provided by utilising the optimal mixing DirAC formulation; which strikes an optimal balance between linear Ambisonic decoding and parametrically enhanced decoding. This approach mitigates spectral artefacts, while simultaneously improving the perceived spatial accuracy. In conclusion, this work suggests that a simple tetrahedral hydrophone array, which can be carried by a single diver, can be utilised in real-time to identify the direction of underwater sound objects with relatively high accuracy; both visually and aurally.

References

[1] Rafaely, B., *Fundamentals of spherical array processing*, volume 8, Springer, 2015.

[2] Zou, N., Swee, C. C., and Chew, B. A., "Vector hydrophone array development and its associated DOA estimation algorithms," in *OCEANS 2006-Asia Pacific*, pp. 1–5, IEEE, 2007.

[3] Farina, A., Armelloni, E., and Chiesi, L., "Experimental evaluation of the performances of a new pressure-velocity 3D probe based on the ambisonics theory," 2011.

[4] Mulayoff, R., Buchris, Y., and Cohen, I., "Differential microphone arrays for the underwate acoustic channel," in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, 2017.

[5] Shipps, J. C. and Abraham, B. M., "The use of vector sensors for underwater port and waterway security," in *Sensors for Industry Conference, 2004. Proceedings the ISA/IEEE*, pp. 41–44, IEEE, 2004.

[6] Moreau, S., Daniel, J., and Bertet, S., "3D sound field recording with higher order ambisonics–Objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, pp. 20–23, 2006.

[7] Gerzon, M. A., "Periphony: With-height sound reproduction," *Journal of the Audio Engineering Society*, 21(1), pp. 2–10, 1973.

[8] Pulkki, V., Delikaris-Manias, S., and Politis, A., *Parametric Time-Frequency Domain Spatial Audio*, John Wiley & Sons, 2017.

[9] Delikaris-Manias, S. and Pulkki, V., "Cross pattern coherence algorithm for spatial filtering applications utilizing microphone arrays," *IEEE Transactions on Audio, Speech, and Language Processing*, 21(11), pp. 2356–2367, 2013.

[10] Nadiri, O. and Rafaely, B., "Localization of multiple speakers under high reverberation using a spherical microphone array and the direct-path dominance test," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 22(10), pp. 1494–1505, 2014.

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

[12] Politis, A., McCormack, L., and Pulkki, V., "Enhancement of ambisonic binaural reproduction using directional audio coding with optimal adaptive mixing," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,*, 2017.

[13] Vilkamo, J., Bäckström, T., and Kuntz, A., "Optimized covariance domain framework for time–frequency processing of spatial audio," *Journal of the Audio Engineering Society*, 61(6), pp. 403–411, 2013.

[14] Williams, E. G., *Fourier acoustics: sound radiation and nearfield acoustical holography*, Academic press, 1999.

[15] Teutsch, H., *Modal array signal processing: principles and applications of acoustic wavefield decomposition*, volume 348, Springer, 2007.

[16] Alon, D. L. and Rafaely, B., "Spatial decomposition by spherical array processing," in *Parametric time-frequency domain spatial audio*, pp. 25–48, Wiley Online Library, 2017.

[17] Ivanic, J. and Ruedenberg, K., "Rotation Matrices for Real Spherical Harmonics. Direct Determination by Recursion," *The Journal of Physical Chemistry A*, 102(45), pp. 9099–9100, 1998.

[18] Delikaris-Manias, S., Vilkamo, J., and Pulkki, V., "Signal-dependent spatial filtering based on

weighted-orthogonal beamformers in the spherical harmonic domain," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(9), pp. 1507–1519, 2016.

[19] Wiggins, B., Paterson-Stephens, I., and Schillebeeckx, P., "The analysis of multi-channel sound reproduction algorithms using HRTF data." in *Audio Engineering Society Conference: 19th International Conference: Surround Sound-Techniques, Technology, and Perception*, Audio Engineering Society, 2001.

[20] Melchior, F., Thiergart, O., Del Galdo, G., de Vries, D., and Brix, S., "Dual radius spherical cardioid microphone arrays for binaural auralization," in *Audio Engineering Society Convention 127*, Audio Engineering Society, 2009.

[21] Shabtai, N. R. and Rafaely, B., "Binaural sound reproduction beamforming using spherical microphone arrays," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 101–105, IEEE, 2013.

[22] Politis, A. and Poirier-Quinot, D., "JSAmbisonics: A Web Audio library for interactive spatial sound processing on the web," 2016.

[23] Borß, C. and Martin, R., "An improved parametric model for perception-based design of virtual acoustics," in *Audio Engineering Society Conference: 35th International Conference: Audio for Games*, Audio Engineering Society, 2009.

[24] Politis, A., "Diffuse-field coherence of sensors with arbitrary directional responses," *arXiv preprint arXiv:1608.07713*, 2016.

[25] Pulkki, V., "Directional audio coding in spatial sound reproduction and stereo upmixing," in *Audio Engineering Society Conference: 28th International Conference: The Future of Audio Technology–Surround and Beyond*, Audio Engineering Society, 2006.

[26] Politis, A., Vilkamo, J., and Pulkki, V., "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, 9(5), pp. 852–866, 2015.

[27] Vilkamo, J. and Backstrom, T., "Time–Frequency Processing: Methods and Tools," in *Parametric Time-Frequency Domain Spatial Audio*, p. 3, John Wiley & Sons, 2017.

[28] McCormack, L., Delikaris-Manias, S., Farina, A., Pinardi, D., and Pulkki, V., "Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localised sub-band sound-field analysis," in *Audio Engineering Society Convention 144*, Audio Engineering Society, 2018.