

Supporting information

Identification of fish sounds in the wild using a set of portable audio-video arrays

Xavier Mouy, Morgan Black, Kieran Cox, Jessica Qualley, Stan Dosso, and Francis Juanes

Contents

1 Assembly and deployment of the audio-video arrays	2
1.1 Large array	2
1.1.1 List of materials and assembly	2
1.1.2 Deployment procedure	6
1.2 Mobile array	6
1.2.1 List of materials and assembly	6
1.2.2 Deployment procedure	6
1.3 Mini array	8
1.3.1 List of materials and assembly	8
1.3.2 Deployment procedure	8
1.4 Considerations for modifying the audio-video arrays	9
1.5 Deployment logs and checklists	10
1.5.1 Pre-deployment checklist	10
1.5.2 Hydrophone position log sheet	11
1.5.3 Deployment log sheet	12
2 Data analysis	13
2.1 Automatic detection of acoustic transients	13
2.2 Time Difference of Arrival	13
2.3 Acoustic localisation	14
2.3.1 Forward model	14
2.3.2 linearised inversion	14
2.3.3 Non-linear inversion	14
3 Optimisation of hydrophone placement	15
4 Localisation of controlled sound sources in the field	16
4.1 Large array	16
4.2 Mini array	16
4.3 Mobile array	19
5 Sound levels measured at Hornby Island and Mill Bay	20

1 Assembly and deployment of the audio-video arrays

This section describes how to assemble and deploy the three audio-video arrays used in this study.

1.1 Large array

1.1.1 List of materials and assembly

Tables 1 and 2 list the material necessary for building the main PVC frame and side camera frame of the large array, respectively. In these tables, the URLs point to websites where the components can be purchased. The placement of the components listed in Tables 1 and 2 is shown in Figs 1 and 2, respectively. The frame is built out of schedule 40 1-1/2" PVC pipes, and 4-way and 3-way PVC Tee fittings. Two galvanized crossed 1/8" wire cables with turnbuckles are used to brace two sides of the array to add rigidity to the frame. Two ropes with galvanized thimbles are attached at the top of the PVC frame to lower or lift the array to/from the seafloor during deployment and recovery. Hydrophones are held snug in a 3/4" 3/4" 1/2" PVC tee fitting that is connected successively to a 1/2" PVC nipple, a 1-1/2" 1/2" PVC bush, and a 1-1/2" 3-way PVC tee fitting cut in half and secured to the array frame using stainless steel hose clamps. While most parts of the frame are snug-fit to allow the array to be disassembled for transportation, some structural components holding the weights are glued with PVC cement to maintain the strength and rigidity of the frame during deployment. Glued components include:

- The four vertical supports of the frame (i.e., E-I-B-I-E-K-C ×4);
- All central components in the middle of the frame holding the acoustic recorder and battery pack (i.e., G-I-G, I-J, I-G, F-J ×2, G-J ×2); and
- Part of the hydrophone mounts (i.e., L1-B ×5, L1-G).

Note that the letters above are defined in the column ID of Tables 1 and 2 and Figs 1 and 2. Gluing the component L1 of the hydrophone mount to the frame allows the position of the hydrophones to remain consistent for each deployment. The remaining part of the hydrophone mount (i.e., L2-L3-L4, Fig. 1c) is typically always fastened to the hydrophone and can be snug-fit into L1 before the deployment. The FishCams are screwed on to the array frame via two 1/2" union fittings (see section 5.3 of Mouy et al., 2020, for a detailed description of the FishCam mounts). The AMAR and its external battery pack are positioned on top of the central horizontal component at the base of the array frame and secured with Smart Band fasteners (Fig. 1e). Hydrophone cables are tied to the frame's PVC pipes with reusable self-gripping Velcro bands. Cables should not be loose to avoid any potential mechanical noise when the system is deployed but should not be so tight that they put strain on connectors or cause excessive bending that could damage them. A video showing the assembly of the large array can be found in the Supporting Information (SuppInfo_Video_LargeArrayAssembly.mp4).

Table 1: List of materials for building the main frame of the large array.

ID	Description	Length (cm)	Quantity	URL
A	Schedule 40 1-1/2" PVC pipe	180	6	https://tinyurl.com/54257y5x
B	Schedule 40 1-1/2" PVC pipe	150	4	https://tinyurl.com/54257y5x
C	Schedule 40 1-1/2" PVC pipe	118	4	https://tinyurl.com/54257y5x
D	Schedule 40 1-1/2" PVC pipe	71	2	https://tinyurl.com/54257y5x
E	Schedule 80 1-1/2" PVC pipe	30	8	https://tinyurl.com/2p9fj9bu
F	Schedule 80 1-1/2" PVC pipe	180	3	https://tinyurl.com/2p9fj9bu
G	Schedule 80 1-1/2" PVC pipe	85	2	https://tinyurl.com/2p9fj9bu
H	Crossed wire cables			
H1	1/4 in. x 5-1/4 in. Stainless Steel Eye and Eye Turn-buckle		4	https://tinyurl.com/pj3ykwuc
H2	Galvanized 1/8" wire cable	240+	4	https://tinyurl.com/yfr69ez5
H3	3/32" x 1/8" zinc-plated clamp		8	https://tinyurl.com/3447ydrw
I	Schedule 40 4-way 1-1/2" PVC Tee fitting		11	https://tinyurl.com/urd6t2hc
J	Schedule 40 1-1/2" slip sling PVC tee		5	https://tinyurl.com/3ftpnff5
K	Schedule 40 45 degree 1-1/2" elbow PVC fitting		4	https://tinyurl.com/mtwr6733
L	Hydrophone mount		6	
L1	Schedule 40 1-1/2" PVC snap tee fitting, snap x socket (or 1-1/2" PVC tee fitting cut in half along the long side)		6	https://tinyurl.com/2p8pwfb5
L2	Schedule 40 1-1/2" to 1/2" reducer bushing fitting, spigot x FPT		6	https://tinyurl.com/8xjy62er
L3	Schedule 40 1/2" PVC nipple NPT		6	https://tinyurl.com/3aj2hheb
L4	Schedule 40 3/4" 3/4" 1/2" PVC tee fitting, socket x socket x FPT		6	https://tinyurl.com/2jt2rdsb
M	1" galvanized thimble attached with rope and stainless steel hose clamps at the top of the PVC frame to lower or lift the array		2	https://tinyurl.com/48z2wxcc
N	Smart Band fasteners		4	https://tinyurl.com/wpphst93
O	Self-gripping Velcro bands			https://tinyurl.com/mapmkkzy
AR	AMAR G3 acoustic recorder		1	https://tinyurl.com/mky5dujh
BP	AMAR external battery pack		1	https://tinyurl.com/mky5dujh
	M36 omnidirectional hydrophones		6	https://tinyurl.com/yc5pfue4
	4-1 hydrophone cable splitter	250	6	https://tinyurl.com/mky5dujh
	Hydrophone mount	250	6	https://tinyurl.com/mky5dujh
C1	Fishcam		1	https://tinyurl.com/mw9axcwe

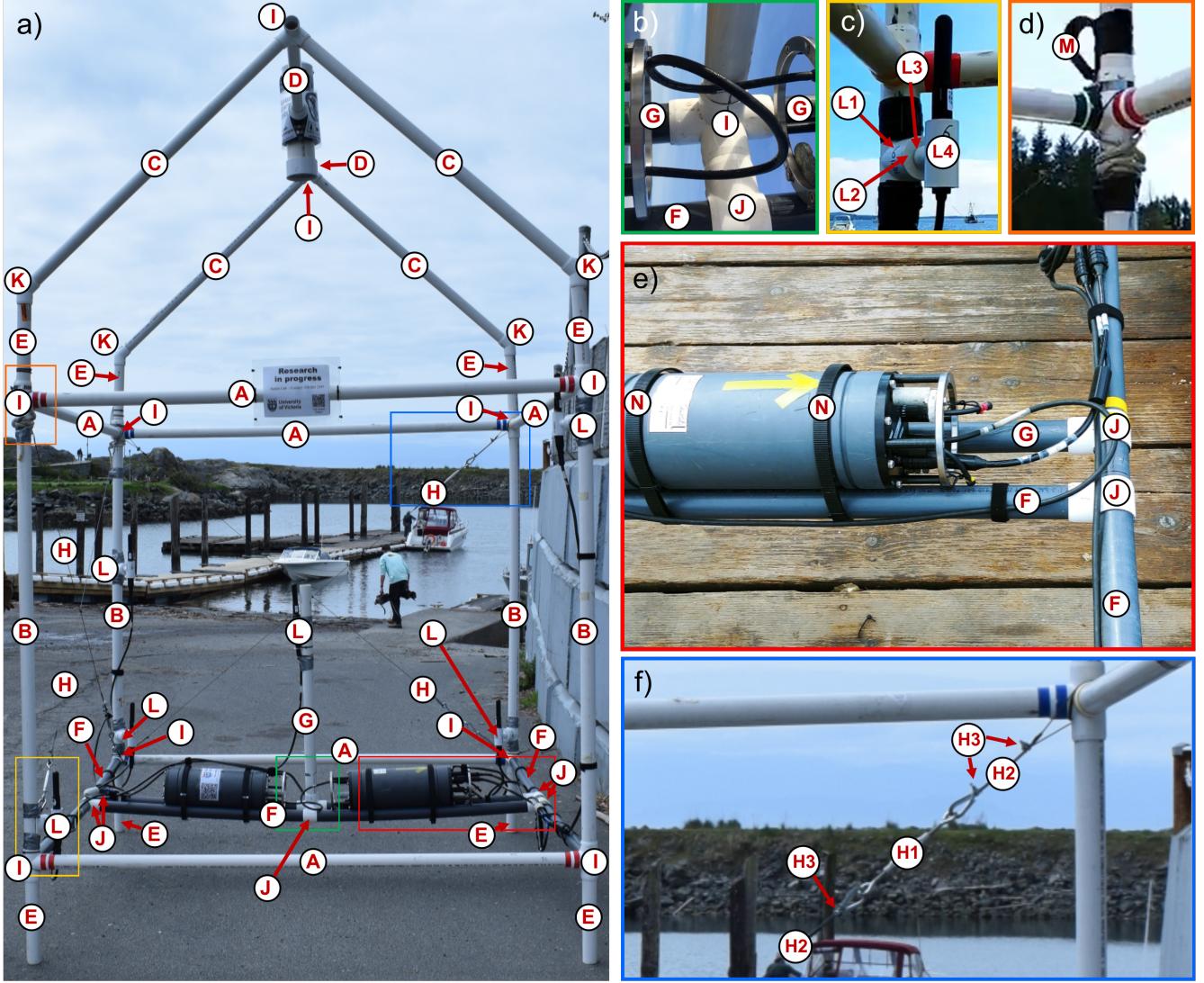


Figure 1: Components of the main frame of the large array. The left panel shows a picture of the entire main frame assembled, while the coloured contoured pictures in the right panel show zoomed-in versions of the components in the coloured boxes (i.e., green, yellow, orange, red and blue) from the left panel. Circled letters correspond to the component identification listed in the first column of Table 1.

Table 2: List of materials for building the side camera frame of the large array.

ID	Description	Length (cm)	Quantity	URL
P	Schedule 40 1-1/2" PVC pipe	93	2	https://tinyurl.com/54257y5x
Q	Schedule 40 1-1/2" PVC pipe	80	2	https://tinyurl.com/54257y5x
R	Schedule 40 1-1/2" PVC pipe	66	1	https://tinyurl.com/54257y5x
S	Schedule 40 1-1/2" PVC pipe	16	2	https://tinyurl.com/54257y5x
L1	Schedule 40 1-1/2" PVC snap tee fitting (or 1 1/2" PVC tee fitting cut in half along the long side)		2	https://tinyurl.com/2p8pwfb5
I	Schedule 40 4-way 1-1/2" PVC Tee fitting		4	https://tinyurl.com/urd6t2hc
C2	Fishcam		1	https://tinyurl.com/mw9axcwe

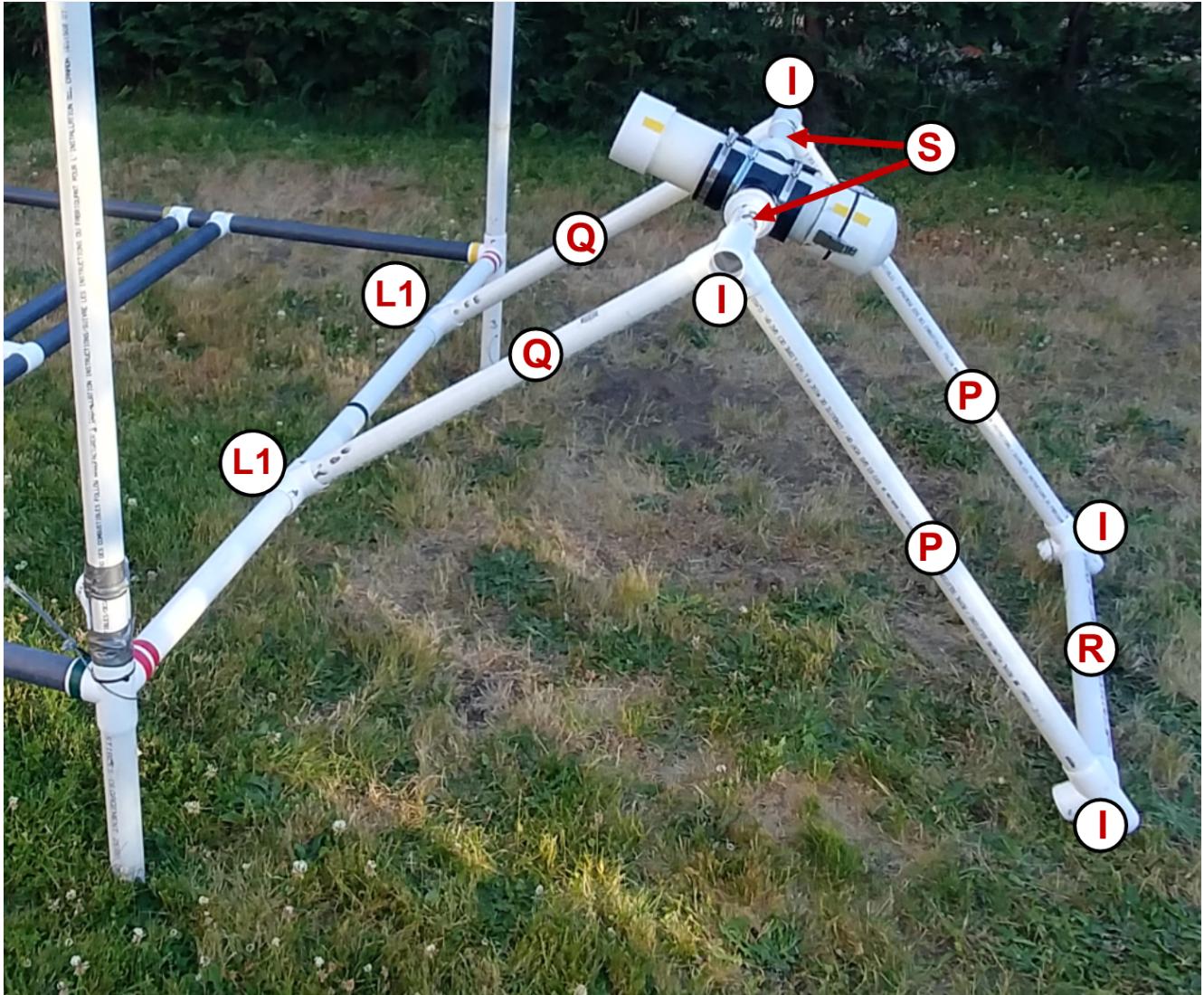


Figure 2: Components of the side camera frame of the large array. Circled letters correspond to the components' identification listed in the first column of Table 2

1.1.2 Deployment procedure

Below is the typical sequence of steps for deploying the large array. A video showing the deployment of the large array can be found in the Supporting Information (SuppInfo_Video_LargeArrayDeployment.mp4).

1. Assemble the array on shore. The side camera frame will not yet be attached to the main frame.
2. Complete the pre-deployment checklist (see section 1.5.1).
3. Start the acquisition on the acoustic recorder and video cameras.
4. Calibrate the hydrophones (e.g., using a piston-phone).
5. Measure the distance between all hydrophones (see section 1.5.2).
6. Produce an acoustic signal for synchronizing the audio and video data (e.g., clap hands in the middle of the array, in the field of view of both cameras).
7. Place the large array on the boat and transit to the deployment location.
8. Place the large array in the water, on the side of the boat, and lower it to the seafloor using two doubled lines going through the lifting rings (i.e., galvanised thimbles) of the array (component M in Fig. 1). Once on the bottom, release one end of each line and bring the lines onboard the boat. At this point, the array is not attached to the boat anymore and is resting on the seafloor. Note that in case of poor water visibility, attach one of the lowering lines to a surface float, which can serve as visual marker during the deployment.
9. Lower weights (e.g., sandbags) to the seafloor beside the large array. Lower the first set of weights using a line and surface float. Lower subsequent weights by clipping them to the line with a carabiner and letting them descend to the seafloor. For our deployments, we used 10 sandbags of 9 kg each.
10. Divers (typically two or three) enter the water and dive to seafloor with the side camera frame.
11. Divers clip the side camera frame to the main frame, attach the weights to each leg of the large array, and detach any of the lines with surface floats.
12. Divers go back onboard the boat and bring surface floats and their lines back onboard.
13. Fill in the deployment log sheet (see section 1.5.3).

1.2 Mobile array

1.2.1 List of materials and assembly

Table 3 lists the material necessary for building the frame for the mobile array. In this table, URLs point to websites where the components can be purchased. Note that the production of the Trident Remotely Operated Vehicle (ROV) is discontinued but can still be found second-hand online via eBay or the Facebook Trident Pilot's Group. Alternatively, other low-priced ROVs, such as the BlueROV from BlueRobotics, can be used. The placement of the components listed in Table 3 is shown in Fig. 3. The frame is made of sections of 3/4" polyethylene pipes connected by polyethylene tee fittings and zip ties (Fig. 3a). The frame is slid from the front side of the Trident ROV (Fig. 3b) and secured by four M3-0.5 screws to the bottom of the Trident (Fig. 3d). Two polyethylene tee fittings are positioned at the bottom-front of the frame to keep the ROV and thrusters out of the sand and mud when the array is resting on the seafloor (Fig. 3e). The acoustic recorder is positioned on top of the ROV and secured to the frame with zip ties. The hydrophones are positioned at the end of each arm of the frame and secured with electrical tape (Fig. 3f). Pieces of buoyant polyethylene foam are attached with zip ties on several sections of the frame to maintain balance and buoyancy. Note that the configuration shown in Fig. 3 is slightly different than the one used for the results in this paper: hydrophone 2 is 8 cm higher, component G is 20 cm longer, and the acoustic recorder is attached on top of the vertical rotor of the Trident. While both configurations work, the one used in the main paper made the mobile array more maneuverable in the field.

1.2.2 Deployment procedure

The mobile array can be deployed by one person. It can be deployed from shore or from a boat. Below is the typical sequence of steps when deploying the mobile array from a boat. A video showing an example of survey performed by the mobile array can be found in the Supporting Information (SuppInfo_Video_MobileArrayDeployment.mp4).

1. Attach the mobile array frame to the ROV with the M3-0.5 screws.
2. Measure the distance between all hydrophones.
3. Start the acquisition on the acoustic recorder.
4. Calibrate the hydrophones (e.g., using a piston-phone).
5. Start recording the video data on the ROV.
6. Produce an acoustic signal for synchronizing the audio and video data (e.g., hand claps in front of the ROV camera).

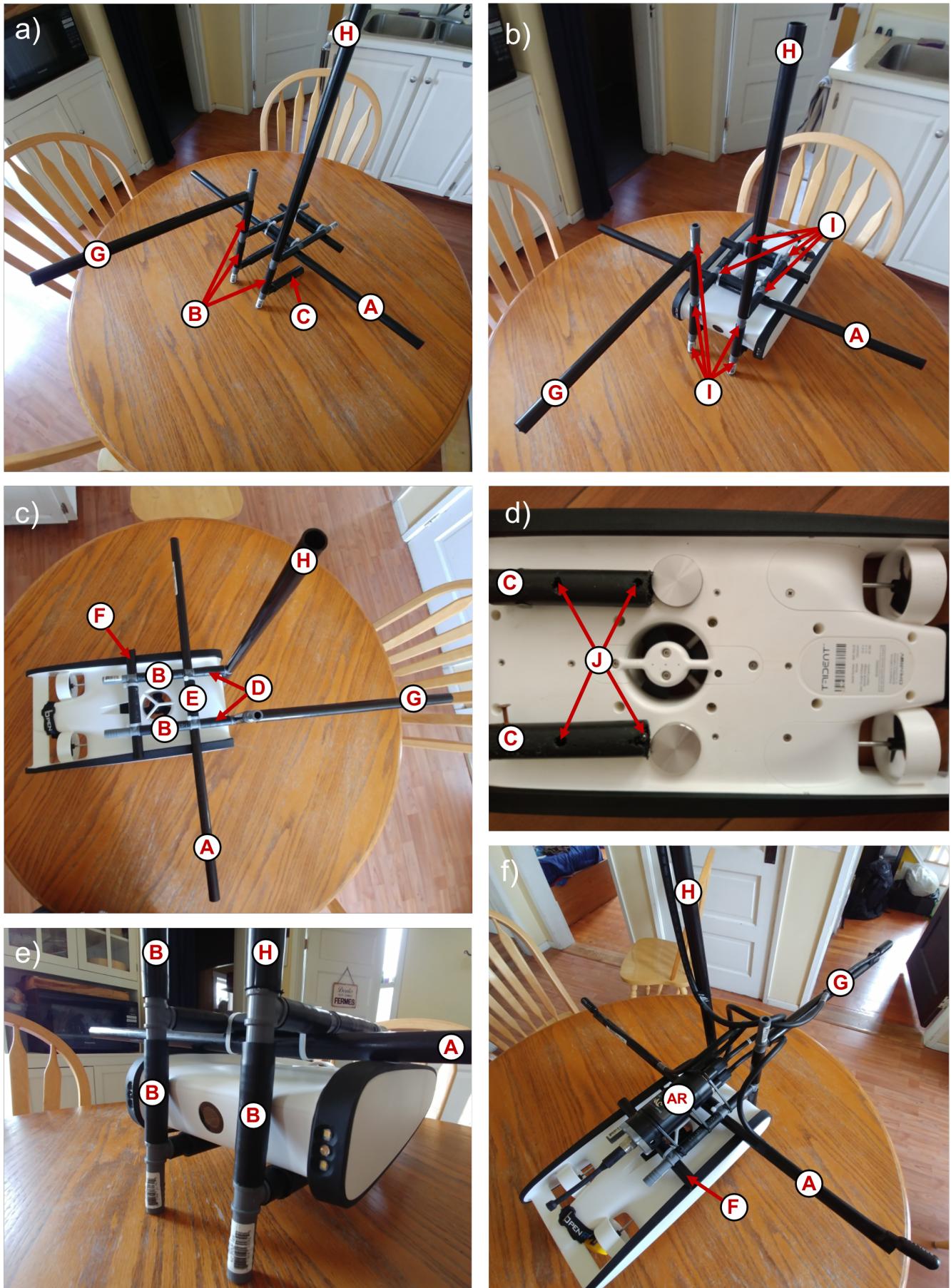


Figure 3: Components of the mobile array. Circled letters correspond to the component identification listed in the first column of Table 3.

Table 3: List of material for building the mobile array.

ID	Description	Length (cm)	Quantity	URL
A	3/4" polyethylene pipe	100	1	https://tinyurl.com/y85rv2e6
B	3/4" polyethylene pipe	8	3	https://tinyurl.com/y85rv2e6
C	3/4" polyethylene pipe	13	2	https://tinyurl.com/y85rv2e6
D	3/4" polyethylene pipe	4	2	https://tinyurl.com/y85rv2e6
E	3/4" polyethylene pipe	6	1	https://tinyurl.com/y85rv2e6
F	3/4" polyethylene pipe	20	1	https://tinyurl.com/y85rv2e6
G	3/4" polyethylene pipe	50	1	https://tinyurl.com/y85rv2e6
H	3/4" polyethylene pipe	60	1	https://tinyurl.com/y85rv2e6
I	3/4" polypropylene insert tee		9	https://tinyurl.com/yuwt6jkj
J	3M-0.5 screws	0.8	4	https://tinyurl.com/yvt3vtsa
AR	SoundTrap ST4300HF		1	https://tinyurl.com/mw7juhn
	Trident ROV		1	https://tinyurl.com/35nzjpzb https://tinyurl.com/2r57xpwd
	HTI-96 omnidirectional hydrophone		4	https://tinyurl.com/yjw2mpyz
	Buoyancy foam		1	https://tinyurl.com/yc7nb39k
	BlueROV alternative to the Trident		1	https://tinyurl.com/2mw4enut

7. Place the mobile array on the boat, go to the deployment location, and anchor the boat.
8. Deploy the mobile array in the water.
9. Pilot the ROV to explore the area to find fish.
10. Land the array on the seafloor or on rocks to quietly record and localise fish sounds.

1.3 Mini array

1.3.1 List of materials and assembly

The mini array is essentially built by combining the side camera frame of the large array (Fig. 2) with the hydrophone polyethylene frame from the mobile array (Fig. 3). The mini array is assembled by using zip ties to attach the mobile array frame on top of the side camera frame of the large array. For a list of materials, see Tables 2 and 3.

1.3.2 Deployment procedure

The mini array can be deployed from a small boat by one person onboard the boat and two divers (Fig. 4). Below is the typical sequence of steps for deploying the mini array. A video showing the deployment of the mini array can be found in the Supporting Information (SuppInfo_Video_MiniArrayDeployment.mp4).

1. Assemble the mini array on shore.
2. Complete the pre-deployment checklist (see section 1.5.1).
3. Start the acquisition on the acoustic recorder and video cameras.
4. Calibrate the hydrophones (e.g., using a piston-phone).
5. Measure the distance between all hydrophones.
6. Produce an acoustic signal for synchronizing the audio and video data (e.g., clap hands in the middle of the array, in the field of view of both cameras).
7. Place the mini array on the boat and transit to the deployment location.
8. Lower weights (e.g., sandbags) to the seafloor with a line tied to a surface float. For our deployment, we used 4 sandbags of 9 kg each.
9. Place the mini array in the water, on the side of the boat.
10. Divers (typically 2) enter the water and dive to seafloor with the mini array.
11. Divers attach the weights to the mini array and detach the lines with the surface float (see step 8).

12. Divers go back onboard the boat or return to shore, depending on the deployment location.
13. Person on the boat brings the surface float and line back onboard.
14. Fill in the deployment log sheet (see section 1.5.3).



Figure 4: Deployment of the mini array from an inflatable boat.

1.4 Considerations for modifying the audio-video arrays

Readers may need to modify these audio-video arrays to fit their particular needs or constraints. Here, we list a few key points that should be considered when modifying the arrays' design:

- **Structure noise.** The array should be put together in such a way that it does not generate noise when subject to significant currents. For PVC frames, this often requires gluing parts of the frame and adding support cables to avoid friction/movement noise. Drilling holes in the structure is also often necessary to avoid air pockets that can increase unwanted sound reflections.
- **Stable and precise hydrophone positions.** Small errors in hydrophone placement can lead to large localisation errors. It is therefore important to have a frame and hydrophone holders that are sturdy enough to not move or bend during deployment. Hydrophones coordinates must be measured as precisely as possible before deployment.
- **Hydrophones placement.** As demonstrated in this study, the placement of the hydrophones impacts the localisation capabilities of the array. It is therefore important to choose hydrophone locations that minimize localisation uncertainties. Using the simulated-annealing approach (and code) we used here should help define the best hydrophone configuration given specific mechanical constraints (i.e., volume available on the frame/platform).

1.5 Deployment logs and checklists

This section provides examples of deployment logs and checklists that can be used when preparing and deploying the arrays. The original files of the checklists and log sheets below can be found on the GitHub repository here: <https://github.com/xaviermouy/XAV-arrays>.

1.5.1 Pre-deployment checklist

Fish Array - Pre-deployment check-list

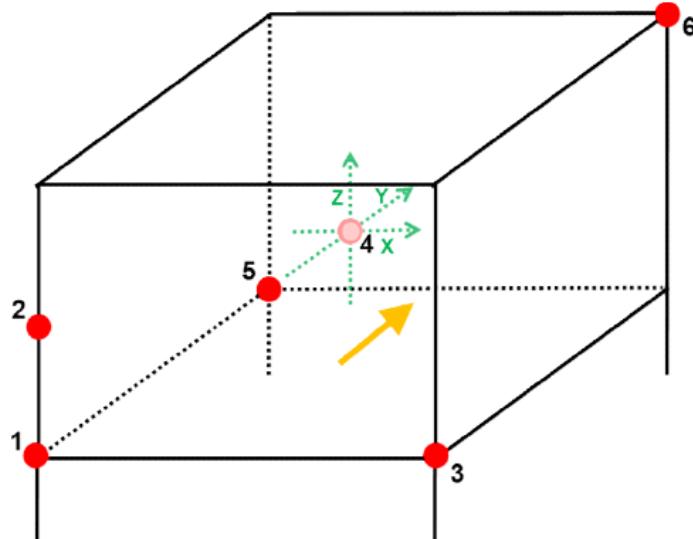
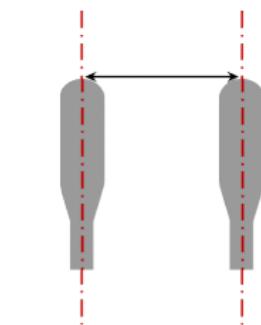
Deployment location:	Deployment ID
Completion date:	Completed by:
Checked	
Camera	
All D-cell batteries are 1.4 V	
Wires properly connected and undamaged	
Camera ribbon connector properly attached/not loose	
Camera lens clean	
Buzzer battery is connected	
Camera time set to UTC	
WittyPi RTC and schedule updated	schedule file name:
Buzzer rang after power turned ON	
Files recording properly (check via SSH)	
Pressure housing clean with no debris	
Plexiglass window cleaned	
O'Ring cleaned and greased with O'Lube	
Acoustic recorder	
AMAR has enough battery for the deployment	Est. time left (days):
AMAR has enough memory for the deployment	
Synchronize AMAR Real Time Clock	
Check duty cycle and sampling frequency (take screenshot)	
All Seacon connectors properly grease and secured	
All dummy plugs are in place	
All nuts and bolts tighten	
GPS	
Clear past waypoints and tracks	
GPS set to track mode	
GPS is in UTC	
Add waypoint for the deployment location	
Enough batteries for the deployment	
Array	
PVC tubes well connected	
Do Not Touch signs attached	
Stainless steal wires properly tighten	
Tie wrap to stabilize stainless steal cable crossing	
Measured distance between each hydrophone	
AMAR well secured to frame	
Hydrophone holders not moving	
Hydrophones properly secured in holders	
Lifting eyes well attached	
mini XAV array: tape legs of array to PVC bar	

1.5.2 Hydrophone position log sheet

Distance between hydrophones - AMAR

Date (UTC, yyyy-mm-dd)
Deployment ID / location
Deployment goals

Measurements	
Hydrophone pair	Distance (cm)
1-2	
1-3	
2-3	
1-5	
1-4	
1-6	
2-4	
2-5	
2-6	
3-4	
3-5	
3-6	
4-5	
4-6	
5-6	
FS01-4	
FS02-4	



1.5.3 Deployment log sheet

Fish Array Deployment Log

Deployment ID				
Target location				
Location name	Lat (d°mm.mmm' N/S)	Lon (d°mm.mmm' E/W)	Depth (m)	
Instruments				
Acoustic recorder (model, S/N)				
Hydrophones (model, S/N)				
Video camera (model, S/N)				
Deployment				
Date (yyyy-mm-dd)		Time zone		
Team/vessel		Weather/sea state		
GPS (model, S/N)		Calibrator (model, S/N)		
Camera - Recording start time (HH:MM:SS)	FC01:	FC02:		
Camera - Time of first buzzer sound (HH:MM:SS)	FC01:	FC02:		
Recorder - Recording start time(HH:MM:SS)				
Recorder - Calibration start time (HH:MM:SS)				
Camera/Recorder - Sync signal description				
Camera/Recorder - Sync signal time (HH:MM:SS)				
Deployment steps	GPS waypoint	Time (HH:MM:SS)	Lat (d°mm.mmm' N/S)	Lon (d°mm.mmm' E/W)
Leaving harbor				
Vessel on site				
Array deployed				
Trident in water				
Trident out of water				
Vessel back to harbor				
Orientation of the array on bottom (degrees from true North)				
Recovery				
Date (yyyy-mm-dd)		Time zone		
Team/vessel		Weather/sea state		
GPS (model, S/N)		Calibrator (model, S/N)		
Camera - Recording stop time (HH:MM:SS)	FC01:	FC02:		
Recorder - Calibration start time (HH:MM:SS)				
Recorder - Recording stop time(HH:MM:SS)				
Camera/Recorder - Sync signal description				
Camera/Recorder - Sync signal time (HH:MM:SS)				
Recovery steps	GPS waypoint	Time (HH:MM:SS)	Lat (d°mm.mmm' N/S)	Lon (d°mm.mmm' E/W)
Leaving harbor				
Vessel on site				
Array retrieved				
Vessel back to harbor				

Notes

2 Data analysis

Data collected by the audio-video arrays are analysed in four steps. First, acoustic transients are automatically detected on the signal from one of the hydrophones (section 2.1). Second, the time-difference of arrivals (TDOAs) between hydrophones of the array are estimated via cross-correlation (section 2.2). Third, the 3D localisation of the detected events is carried out (section 2.3) using linearised inversion (for the large array) or fully non-linear inversion (for the mini and mobile arrays). Finally, 3D acoustic localisations are matched with the video recordings to identify the species and behaviour of the fish that emitted the sound. This section provides further details for each of these steps.

2.1 Automatic detection of acoustic transients

All acoustic events (including, but not limited to fish sounds) are automatically detected on data from one of the hydrophones of the array (hydrophone 4 for the large array, and hydrophone 2 for the mini and mobile arrays). The detection is performed by calculating and denoising the spectrogram, then by segmenting the denoised spectrogram to identify acoustic transients.

The spectrogram is calculated using 0.0625 s frames with time steps of 0.01 s and 0.0853 s FFTs. Given that most fish sounds reported in the literature have a peak frequency below 5 kHz (Kasumyan, 2008, 2009), the spectrogram is truncated to only keep frequencies from 0 to 6 kHz. Magnitude values are squared to obtain energy and expressed in decibels. To improve the signal-to-noise ratio of fish sounds and attenuate tonal sounds from vessels, the spectrogram is equalized using a median filter, calculated with a sliding window, for each row (frequency) of the spectrogram. The equalized spectrogram, $\hat{S}[t, f]$, at each time bin, t , and frequency bin, f , is calculated as

$$\hat{S}[t, f] = S[t, f] - S_{med}[t, f], \quad (1)$$

where $S[t, f]$ is the original spectrogram and $S_{med}[t, f]$ is the median spectrogram calculated as:

$$S_{med}[t, f] = \text{median}(S[t - k, f], S[t - k + 1, f], \dots, S[t, f], \dots, S[t + k - 1, f], S[t + k, f]), \quad (2)$$

where the median is calculated on a window centred on the t^{th} sample and has a duration of $2k + 1$ bins. Here, we choose a median window equivalent to a 3 s duration ($k = 150$), which removes constant tonal components from vessels without removing the longer grunting sounds from fish.

Once the spectrogram is calculated and equalized, it is segmented by calculating the local energy variance on a two-dimensional (2D) kernel of size $\Delta T \times \Delta F$. The resulting matrix, S_{var} , is defined as

$$S_{var}[t, f] = \frac{1}{(\Delta T \Delta F) - 1} \sum_{i=t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} \sum_{j=f-\frac{\Delta F}{2}}^{f+\frac{\Delta F}{2}} |\hat{S}[i, j] - \mu|^2, \quad (3)$$

where μ is the mean over the 2D kernel

$$\mu = \frac{1}{(\Delta T \Delta F)} \sum_{i=t-\frac{\Delta T}{2}}^{t+\frac{\Delta T}{2}} \sum_{j=f-\frac{\Delta F}{2}}^{f+\frac{\Delta F}{2}} \hat{S}[i, j]. \quad (4)$$

The number of time and frequency bins of the kernel are chosen to be equivalent to 0.1 s and 300 Hz, respectively. Bins of the spectrogram with a local variance less than 20 were set to zero and all the other bins were set to one. Bounding boxes of contiguous bins in the binarized spectrogram are then defined using the outer border following algorithms described in Suzuki and Be (1985). These bounding boxes define the start and stop times and the minimum and maximum frequencies of the acoustic events to localise.

2.2 Time Difference of Arrival

The TDOAs are obtained by cross-correlating acoustic events detected on the recording from the reference-hydrophone with the recordings from the other hydrophones. Before performing the cross-correlation, each recording is band-pass filtered in the frequency band determined by the detector using an eighth-order, zero-phase, forward-backward Butterworth filter (Gustafsson, 1996), and up-sampled using the FFT method (Oppenheim & Schafer, 2009), to obtain a time resolution of 0.1 μ s. The TDOA of a detected signal between a pair of hydrophones is defined as the lag time of the maximum peak in the normalized cross-correlation within the range of possible values for the array (i.e., ± 1.9 ms for the large array, ± 0.8 ms for the mini array, and ± 0.6 ms for the mobile array). Only TDOAs with a peak correlation amplitude greater than 0.5 are considered for localisation.

2.3 Acoustic localisation

2.3.1 Forward model

Given that fish sounds only propagate over small distances (Amorim, 2015) and that hydrophones of the audio-video arrays are separated by less than 3 m, it is assumed, for the localisation, that the effects of refraction are negligible and that the sound velocity, v , is constant (here, $v = 1484$ m/s). In such case, the TDOA Δt_{ij} between hydrophones i and j is defined by

$$\Delta t_{ij} = \frac{1}{v} \left(\sqrt{(X - x_i)^2 + (Y - y_i)^2 + (Z - z_i)^2} - \sqrt{(X - x_j)^2 + (Y - y_j)^2 + (Z - z_j)^2} \right), \quad (5)$$

where x, y, z are the known 3D Cartesian coordinates of hydrophones i and j relative to the array centre, and X, Y, Z are the unknown coordinates of the acoustic source ($M = 3$ unknowns).

Localising the acoustic source is a non-linear problem defined by

$$[d_k; k = 1, N]^T = \mathbf{d} = \mathbf{d}(\mathbf{m}), \quad (6)$$

where \mathbf{d} represents the N measured TDOA data and $\mathbf{d}(\mathbf{m})$ the modelled TDOA data with $\mathbf{m} = [X, Y, Z]^T$ (in the common convention adopted here bold lower-case symbols represent vectors and bold upper-case symbols represent matrices). The large array has 6 hydrophones and can provide 5 independent TDOAs measurements ($N = 5$), while the mini and mobile arrays have 4 hydrophones which provide 3 independent TDOA measurements ($N = 3$).

2.3.2 linearised inversion

Acoustic localisation for the large array is performed using linearised inversion. The localisation problem in Eq. 6 can be linearised by starting at a location \mathbf{m}_0 , and iteratively solving for small perturbations $\delta\mathbf{m}$, to the model \mathbf{m} (Mouy et al., 2018). Assuming errors on the data are identical and independently Gaussian-distributed random variables, the maximum-likelihood solution is

$$\delta\mathbf{m} = [\mathbf{A}^T \mathbf{A}]^{-1} \mathbf{A}^T \delta\mathbf{d}, \quad (7)$$

where $\delta\mathbf{d} = \mathbf{d} - \mathbf{d}(\mathbf{m}_0)$, and \mathbf{A} is the $N \times M$ Jacobian matrix of partial derivatives with elements

$$A_{ij} = \frac{\partial d_i(\mathbf{m}_0)}{\partial m_j}; \quad i = 1, \dots, N; \quad j = 1, \dots, M. \quad (8)$$

The location \mathbf{m} of the acoustic source can be estimated by solving for $\delta\mathbf{m}$ and redefining iteratively

$$\mathbf{m}_{l+1} = \mathbf{m}_l + \alpha \delta\mathbf{m}; \quad l = 0, \dots, L; \quad 0 < \alpha \leq 1, \quad (9)$$

until convergence (i.e., until $\|\mathbf{m}_{l+1} - \mathbf{m}_l\|_2 < 0.01$ m). In Eq. 9, α is a step-size damping factor (here, $\alpha = 0.1$) and L is the maximum number of iterations (here, $L = 100$). The localisation uncertainties, about the final solution, are estimated from the diagonal elements of the model covariance matrix \mathbf{C}_m defined by

$$\mathbf{C}_m = \sigma^2 [\mathbf{A}^T \mathbf{A}]^{-1}, \quad (10)$$

where σ^2 is the variance of the TDOA measurement errors. The localisation process is repeated for each detected sound by choosing five different starting models \mathbf{m}_0 . One is always selected at the centre of the array (i.e., $\mathbf{m}_0 = [0, 0, 0]^T$), and the four others are randomly drawn from a uniform distribution within the volume of the array. The localisation result with the smallest data misfit is considered the best solution. The variance of the TDOA measurement errors, σ^2 in Eq. 10, is estimated for each localisation by

$$\sigma^2 = \frac{1}{N-3} \sum_{i=1}^N (d_i - d_i(\mathbf{m}))^2. \quad (11)$$

2.3.3 Non-linear inversion

The spacing between hydrophones for the mini and mobile arrays is smaller than for the large array and most sound sources to localise are outside the volume of the array. In such cases the problem becomes highly non-linear and localising using linearised inversion is less appropriate, as it has difficulties converging towards a final solution and the linearised uncertainty estimates become unreliable. Consequently, acoustic localisation for the mini and mobile arrays is performed using a fully

non-linear inversion. A 3D spatial search grid of pre-computed time-difference data, $\mathbf{d}(\mathbf{m})$ is defined using Eq. 5 and simulated sound sources positioned every 2 cm from -3 m to 3 m in the x and y axes, and from -1 m to 3 m in the z axis. Assuming errors in the data are identical and independently Gaussian distributed, the un-normalized likelihood $L(\mathbf{m}, \sigma)$ is estimated at each location \mathbf{m} of the search grid as

$$L(\mathbf{m}, \sigma) = \exp \left[-\frac{1}{2\sigma^2} |\mathbf{d} - \mathbf{d}(\mathbf{m})|^2 \right], \quad (12)$$

where $|\mathbf{d} - \mathbf{d}(\mathbf{m})|^2 = \sum_{i=1}^N (d_i - d_i(\mathbf{m}))^2$, and σ^2 is the variance of the TDOA measurement errors estimated at the grid location \mathbf{m}_{ML} that minimizes $|\mathbf{d} - \mathbf{d}(\mathbf{m})|^2$. The likelihood values of the $M_x \times M_y \times M_z$ grid are then normalized by $S = \sum_{i=1}^{M_x} \sum_{j=1}^{M_y} \sum_{k=1}^{M_z} L(x_i, y_j, z_k)$ to obtain a normalized posterior probability density in 3D. The location \mathbf{m}_{ML} has the maximum likelihood and is considered as the estimate of the sound source location. Uncertainties of the localisation are then characterised as the limits of the 68% credibility interval of the marginal probability distributions, P_x , P_y , and P_z defined for each axis, x , y , and z of the 3D grid, respectively. As the mini and mobile array can only provide 3 TDOAs, σ^2 is not updated for each localisation, as for the large array, but for groups of 3 or more consecutive localisations and assumed constant over the recording analysed.

3 Optimisation of hydrophone placement

For the large array, the placement of the hydrophones was defined so as to minimize the overall localisation uncertainty. This was achieved using the simulated annealing optimisation algorithm (Kirkpatrick et al., 1983) and followed the procedure developed in Dosso and Sotirin (1999). The optimisation consisted of finding the x , y , and z coordinates of the six hydrophones (18 parameters) that minimizes the average localisation uncertainty of 600 simulated sound sources placed on a 2 m radius sphere around the centre of the array. For a given hydrophone configuration \mathbf{m} , the average localisation uncertainty, $E(\mathbf{m})$, is defined as

$$E(\mathbf{m}) = \sum_{s=1}^S \sqrt{\sigma_{s_x}^2 + \sigma_{s_y}^2 + \sigma_{s_z}^2} / S, \quad (13)$$

where S is the number of simulated sound sources ($S = 600$), and $\sigma_{s_x}^2$, $\sigma_{s_y}^2$, and $\sigma_{s_z}^2$ are the variances of the localisation uncertainties of the simulated sound source s in the x , y , and z directions, respectively, and corresponding to the diagonal elements of the covariance matrix defined in Eq. 10. Note that in this section, the set of unknown model parameters \mathbf{m} corresponds to the hydrophone positions, and not to the sound source location as in section 2.3. The simulated annealing process starts by randomly choosing the initial hydrophone positions, \mathbf{m} , within the defined bounds and calculating the corresponding $E(\mathbf{m})$. Here, we constrain the hydrophones to be within the interval $[-1, 1]$ m in each dimension. Then, an updated hydrophone placement \mathbf{m}' , is proposed by perturbing one of the components of \mathbf{m} (i.e. one coordinate for one hydrophone) and calculating the associated $E(\mathbf{m}')$. If $\Delta E = E(\mathbf{m}') - E(\mathbf{m}) \leq 0$, then the proposed hydrophone placement is accepted and \mathbf{m} is updated to \mathbf{m}' . If $\Delta E > 0$, then a value $\xi \in [0, 1]$ is drawn from a uniform random distribution and an acceptance probability P is defined as

$$P = e^{-\Delta E/T}, \quad (14)$$

where T is a control parameter of the simulated annealing algorithm called the annealing temperature. If $\xi \leq P$, then the proposed hydrophone placement is accepted and \mathbf{m} is updated to \mathbf{m}' . If $\xi > P$, then the proposed hydrophone placement is rejected and \mathbf{m} is not updated. The simulated annealing process repeats these steps many times, cycling through the parameters, until convergence. The initial value of T is set high enough to accept at least 80% of the proposed hydrophone placements and thoroughly explore the search space (melting phase). The annealing temperature is then reduced by 10% every 100 iterations until only 0.1% of the proposed hydrophone placements are accepted (freezing phase). The proposed \mathbf{m}' are defined by perturbing one parameter m_i of \mathbf{m} individually at each iteration such that

$$m_i' = m_i + \eta \Delta_i, \quad (15)$$

where η is randomly drawn from a normal distribution of mean zero and standard deviation 0.125, and Δ_i is the width of the search interval for that parameter. Here, the search interval for each hydrophone coordinate is $[-1, 1]$ m, so $\Delta_i = 2$ m.

Fig. 5 depicts the evolution of hydrophone coordinates during the simulated annealing search process. For about the first half of the iterations (i.e., at higher annealing temperatures), the full range of allowed coordinate values for each hydrophone were explored (i.e., $[-1, 1]$). During the second half of the iterations (i.e., smaller annealing temperatures), each parameter converged progressively towards its final value. At the end of the simulated annealing process, five hydrophones

converged to a corner of the search space (i.e., cubic volume), and one converged to the centre of the array. The simulated annealing process was repeated five times (using a different initial hydrophone placement) and always led to the same relative hydrophone geometry (with possible rotations around the centre of the array). For practical reasons (i.e., restriction in cable length), the hydrophone placement used for the large array is a rotated version of the hydrophone placement shown in Fig. 5. An animation showing the simulated annealing process can be found in the Supporting Information (SuppInfo_Video_SimulatedAnnealingOptimisation.mp4).

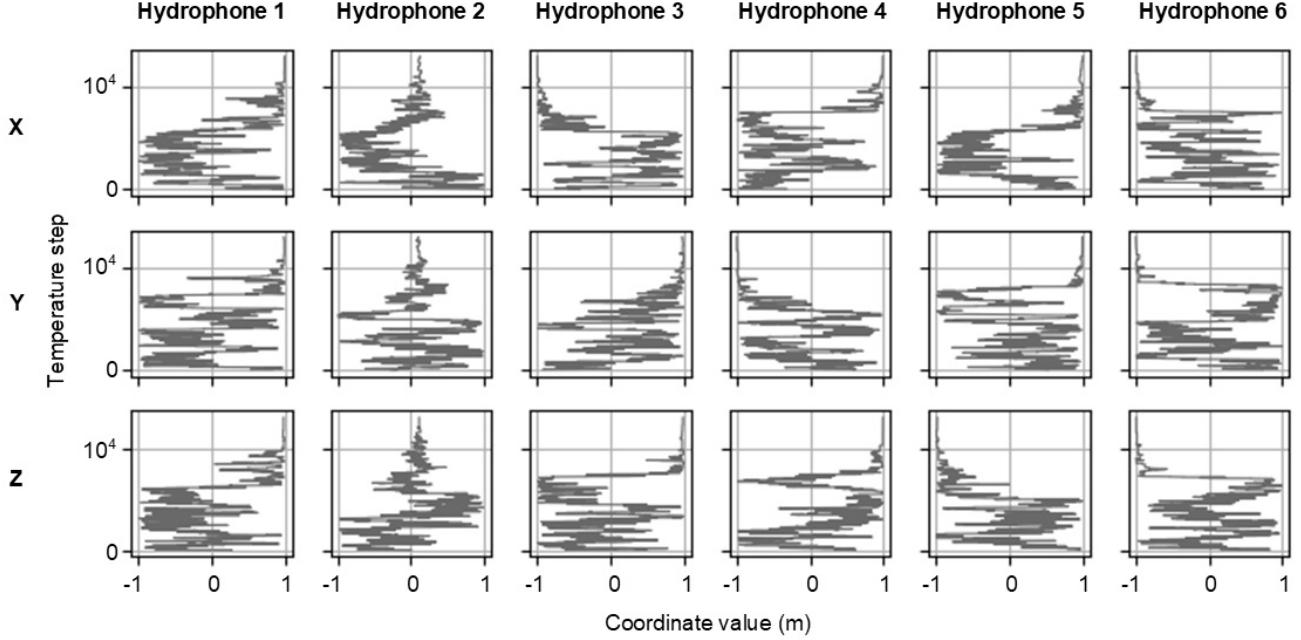


Figure 5: Optimisation of hydrophone placement using simulated annealing: X (top row), Y (middle row), and Z (bottom row) coordinates of each hydrophone (columns) at each iteration of the simulated annealing process.

4 Localisation of controlled sound sources in the field

For each array, the accuracy of the acoustic localisation was tested in the field using a source emitting known sounds at known locations.

4.1 Large array

For the large and mini arrays, the Trident underwater ROV was used as the controlled sound source. The characteristic noise from the ROV's thrusters was received by all hydrophones and was used for localisation. The accuracy of the localisation was then assessed when the ROV was within the field of view of the cameras. Fig. 6 shows the acoustic localisation of the underwater ROV over 6 s as it approaches the large array. The broadband sounds from the ROV's thrusters (Fig. 6a) were localised on the left side of the array at approximately 1 m from the seafloor and with estimated localisation uncertainties less than 20 cm in each dimension (Fig. 6b,c). While the precise coordinates of the ROV were not manually measured *in situ*, these acoustic localisations corresponded very closely to the ROV positions and movements observed from video camera C2 (i.e., approaching the array from the left side at constant depth, then facing hydrophone 4, see Fig. 6d,e).

4.2 Mini array

Fig. 7 shows the acoustic localisation of the underwater ROV over 6 s as it faces the mini array. The broadband sounds from the ROV's thrusters (Fig. 7a) were localised in front of the mini array at approximately 1.5 m above the seafloor and localisation uncertainties were very small along all axes (i.e., less than 4 cm). While the precise coordinates of the ROV were not manually measured *in situ*, these acoustic localisations corresponded very closely to the ROV positions and movements observed from the video camera (i.e., ROV approximately 2 m in front of the array, facing the camera, then slowly moving backwards and up).

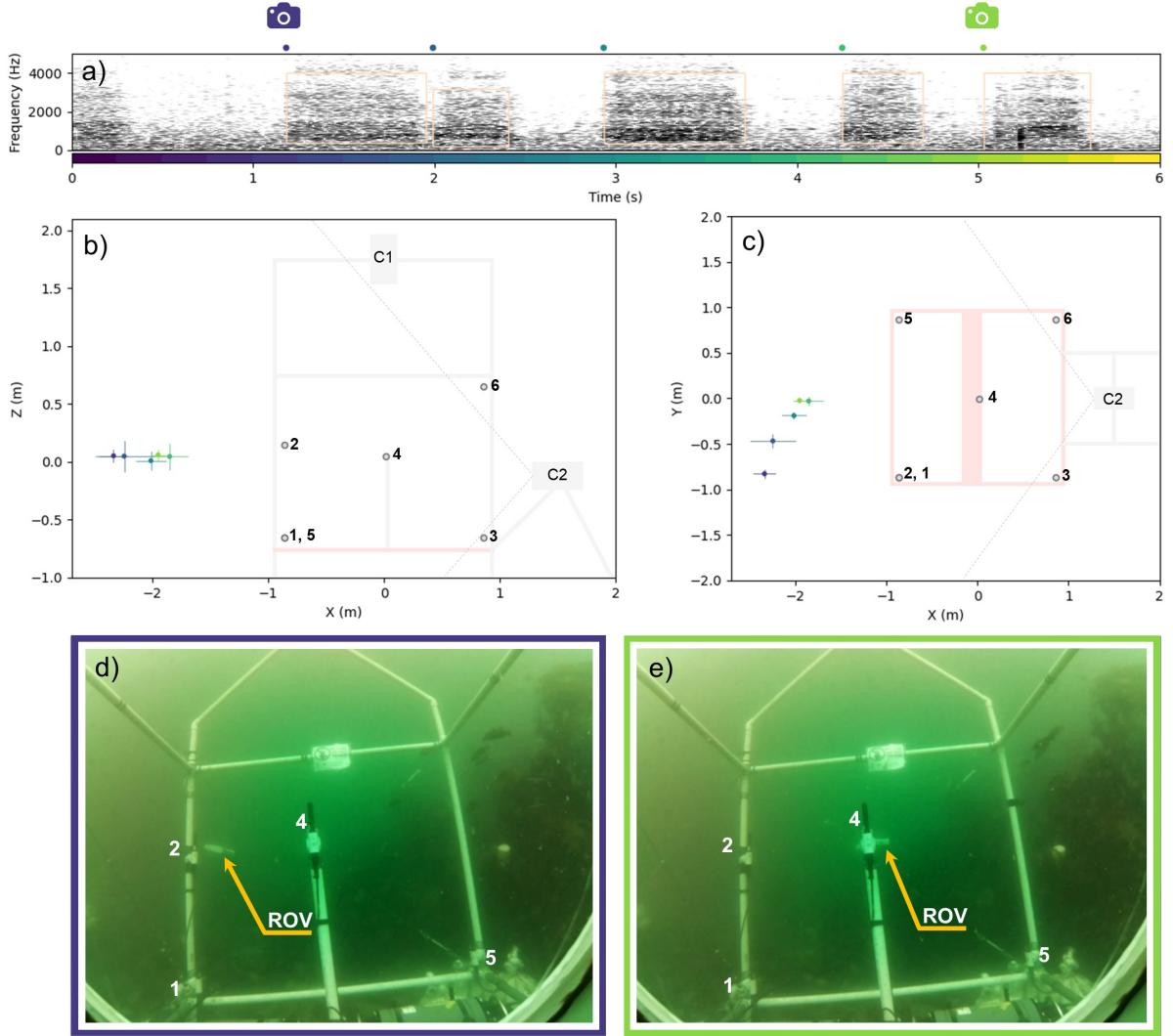


Figure 6: Acoustic localisation of the underwater ROV using the large array deployed at Ogden Point (17 Jun. 2019). (a) Spectrogram of the acoustic recording acquired by hydrophone 4 (frame: 0.0624 s, FFT: 0.0853 s, step size: 0.01 s, Hanning window). Beige boxes indicate the time and frequency limits of the sounds that were automatically detected. Dots at the top of the spectrogram indicate the colours associated to the start time of each detection (see colour scale on the x -axis) and used for the localisation. Coloured camera icons indicate the time of the camera frames shown in panels (d) and (e). (b) Side and (c) top view of the large array. Dashed grey lines in panels (b) and (c) indicate the field of view of camera C2. Coloured dots and lines represent the coordinates and uncertainty (standard deviation) of the acoustic localisations, respectively. (d) Image taken by video camera C2 at $t = 1.2$ s, and (e) $t = 5$ s, showing the underwater ROV approaching the array. Numbers in panels (b), (c), (d), and (e) correspond to the hydrophone identification numbers.

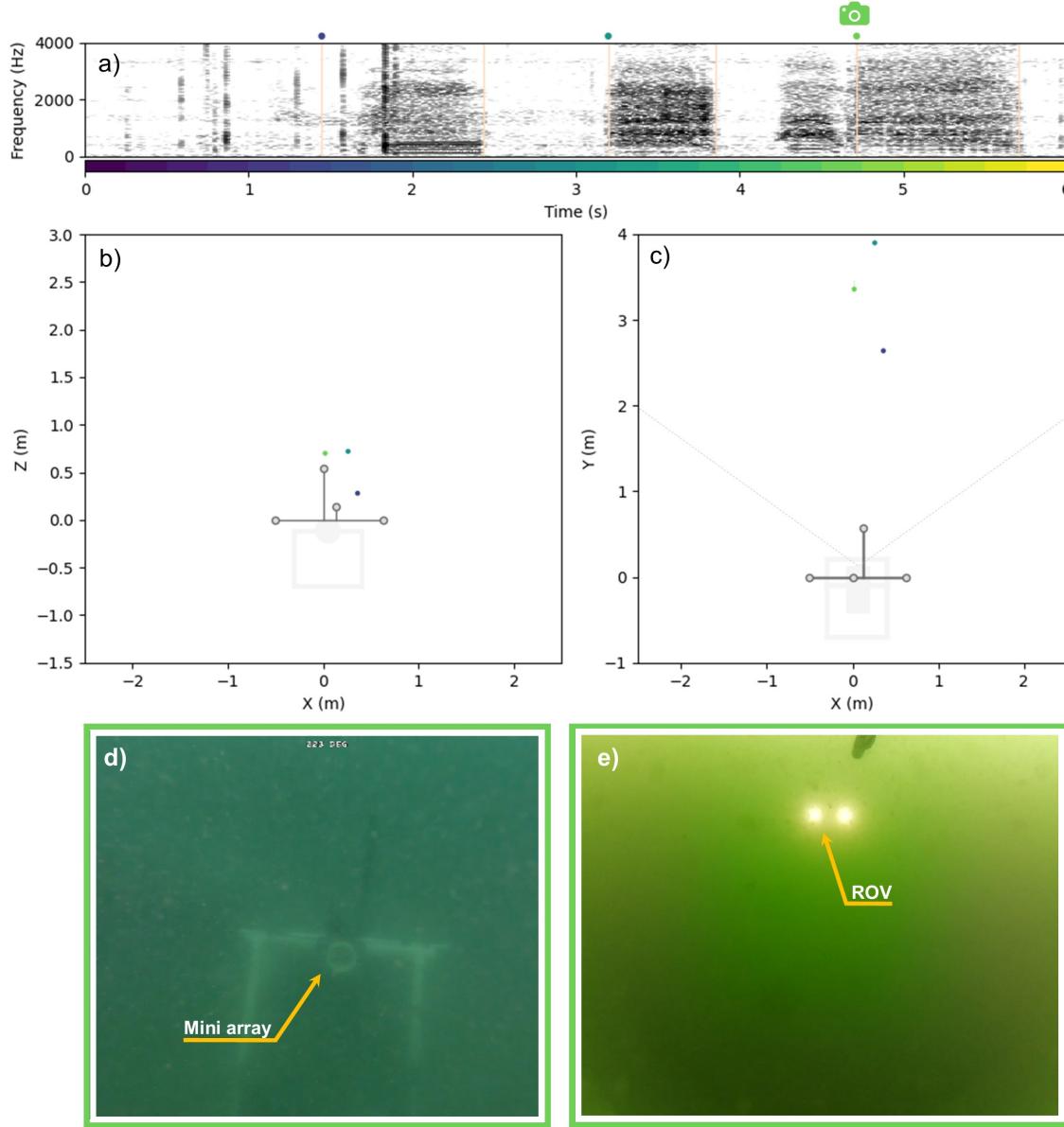


Figure 7: Acoustic localisation of the underwater ROV using the mini array deployed at Mill Bay (1 Aug. 2019). (a) Spectrogram of the acoustic recording acquired by hydrophone 2 (frame: 0.0624 s, FFT: 0.0853 s, step size: 0.01 s, Hanning window). Beige boxes indicate the time and frequency limits of the underwater ROV sounds that were automatically detected. Dots at the top of the spectrogram indicate the colours associated to the start time of each detection (see colour scale on the x -axis) and used for the localisation. The green camera icon indicates the time of the camera frames showed in panels (d) and (e). (b) Rear and (c) top view of the mini array. Dashed grey lines in panel (c) indicate the field of view of the camera. Coloured dots and lines represent the coordinates and uncertainty (68% credibility interval) of the acoustic localisations, respectively. (d) Image taken by the underwater ROV at $t = 4.7$ s showing the mini array. (e) Image taken from the video camera of the mini array at $t = 4.7$ s, showing the underwater ROV in front of the array.

4.3 Mobile array

For the mobile array, the acoustic localisation was tested by emitting sounds from an acoustic projector. A Lubel Labs LL96 acoustic projector was connected to a Clarion Marine APX280M power amplifier and an Apple Ipod audio player, and was deployed on the seafloor, in 3 m of water, off a marina dock at Macaulay Point. The mobile array was also deployed on the seafloor and was located approximately 1 m in front of the projector. The acoustic projector was set to play a series of fish sounds that were then localised by the mobile array. To test the localisation in different directions, measurements were repeated with the mobile array oriented at different angles from the acoustic projector (0, 90, 180, and 270°). The orientation of the mobile array and its distance from the source were adjusted using the compass readings from the ROV and visually from the surface. Fig. 8 shows the localisation results from this experiment. The localisation from the mobile array can resolve the source bearing and elevation angles relatively well, but is much less accurate at estimating range, particularly for sound sources located on either sides of the array (i.e., 270 and 90 °, blue and green dots in Fig. 8, respectively).

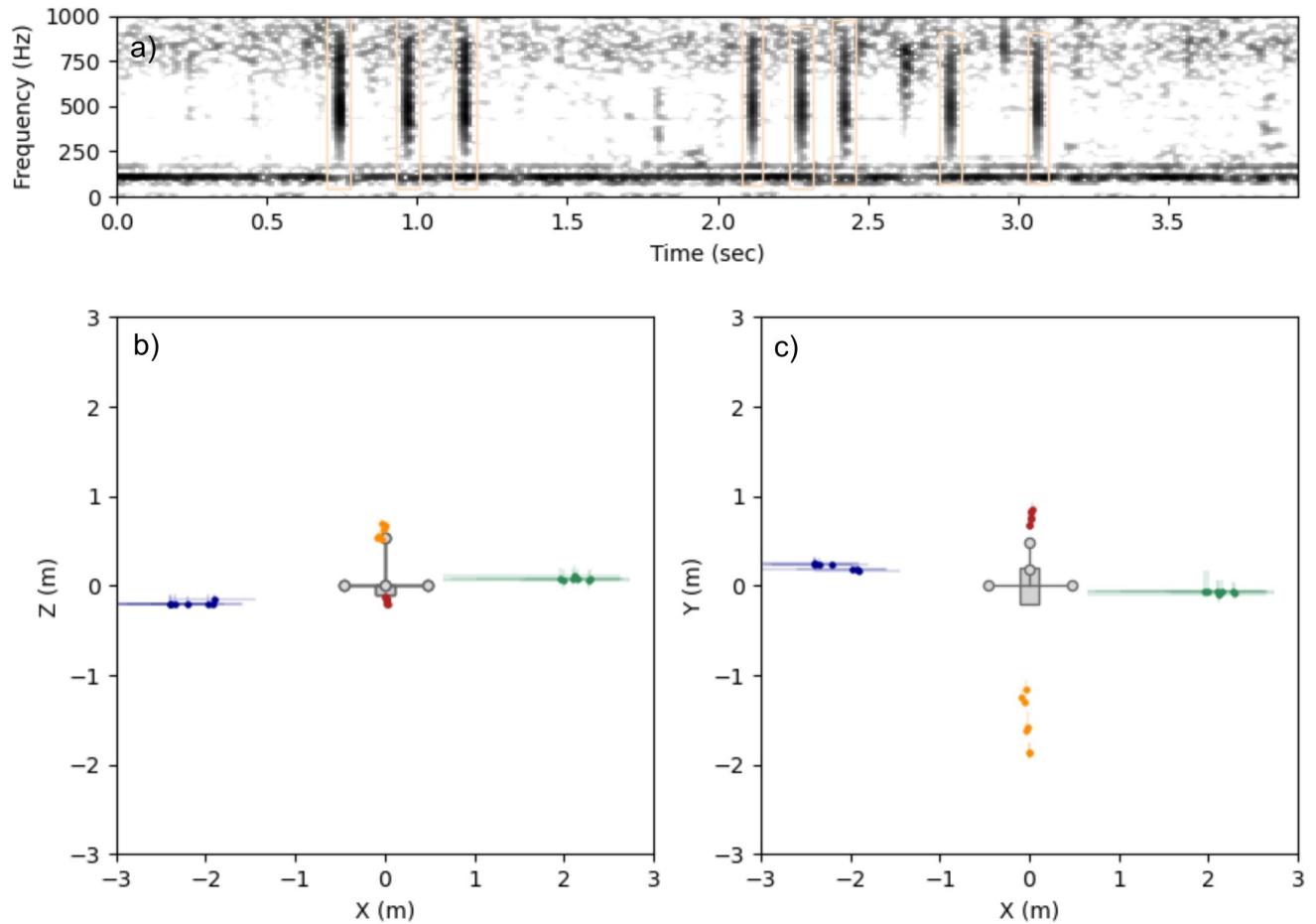


Figure 8: Localisation of an acoustic projector using the mobile array deployed at Macaulay Point (10 Sep. 2020). (a) Spectrogram of the acoustic recording acquired by hydrophone 2 (frame: 0.0624 s, FFT: 0.0853 s, step size: 0.01 s, Hanning window). Beige boxes indicate the time and frequency limits of the fish sounds that were emitted by the acoustic projector and automatically detected. (b) Rear and (c) top view of the mobile array. Coloured dots and lines represent the coordinates and uncertainty (68% credibility interval) of the acoustic localisations, respectively. Red, green, orange, and blue correspond to acoustic localisations when the acoustic projector was located approximately at coordinates (0,1,0), (1,0,0), (0,−1,0), and (−1,0,0), respectively.

5 Sound levels measured at Hornby Island and Mill Bay

This section describes the noise levels that were used to estimate the detection range of lingcod and quillback rockfish sounds. Figs 9 and 10 show the power spectral density (PSD) probability and the long-term spectrogram of the measured sound levels at Mill Bay, respectively. Figs 11 and 12 show the power spectral density (PSD) probability and the long-term spectrogram of the measured sound levels at Hornby Island, respectively.

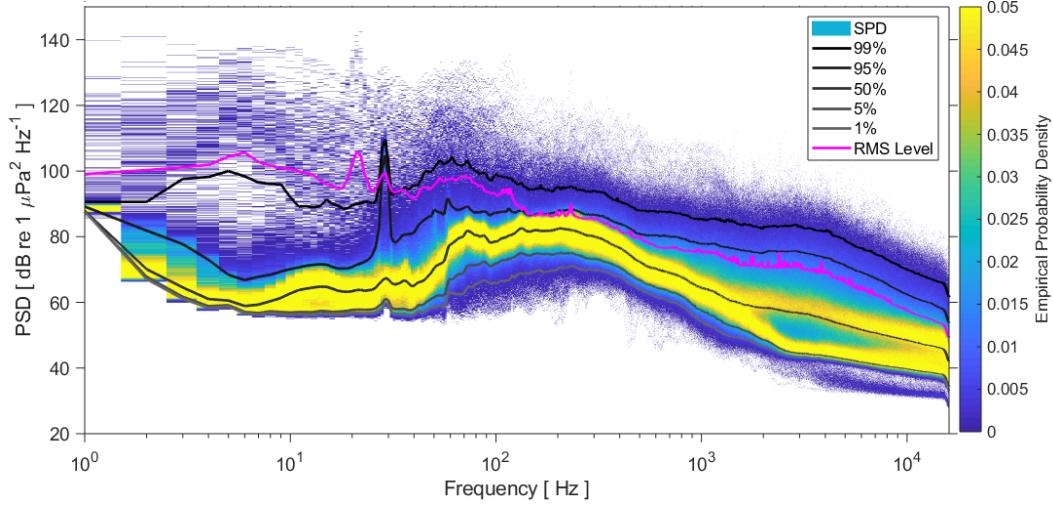


Figure 9: Power spectral density (PSD) probability of sound levels measured at Mill Bay from 18 Aug. to 2 Sep. 2019 (FFT size: 32,0000 samples, overlap: 50%, window function: Hanning). Solid gray and black lines indicate the percentile levels, and the pink solid line indicates RMS levels.

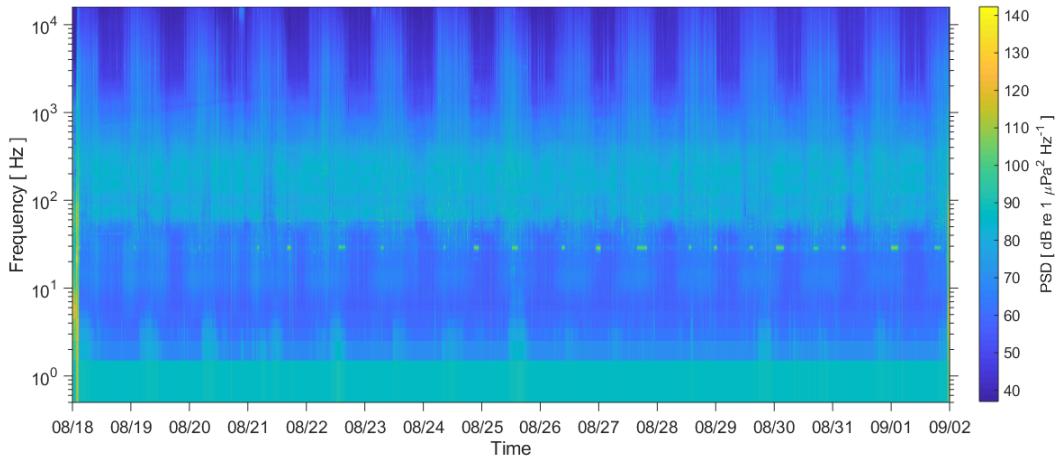


Figure 10: Long-term spectrogram of sound levels measured at Mill Bay from 18 Aug. to 2 Sep. 2019 (FFT size: 32,0000 samples, overlap: 50%, window function: Hanning).

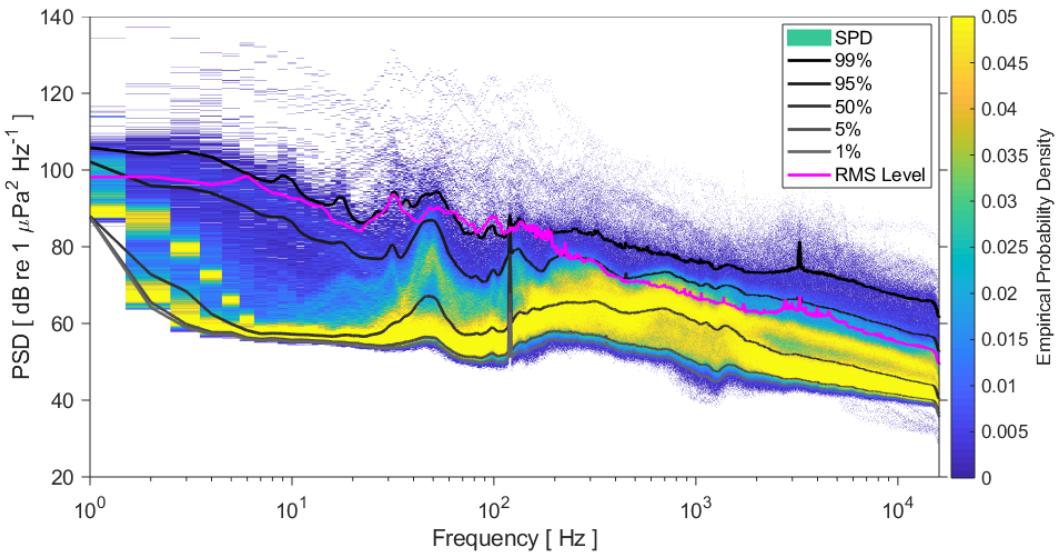


Figure 11: Power spectral density (PSD) probability of sound levels measured at Hornby Island from 15 Sep. to 22 Sep. 2019 (FFT size: 32,0000 samples, overlap: 50%, window function: Hanning). Solid gray and black lines indicate the percentile levels, and the pink solid line indicates RMS levels.

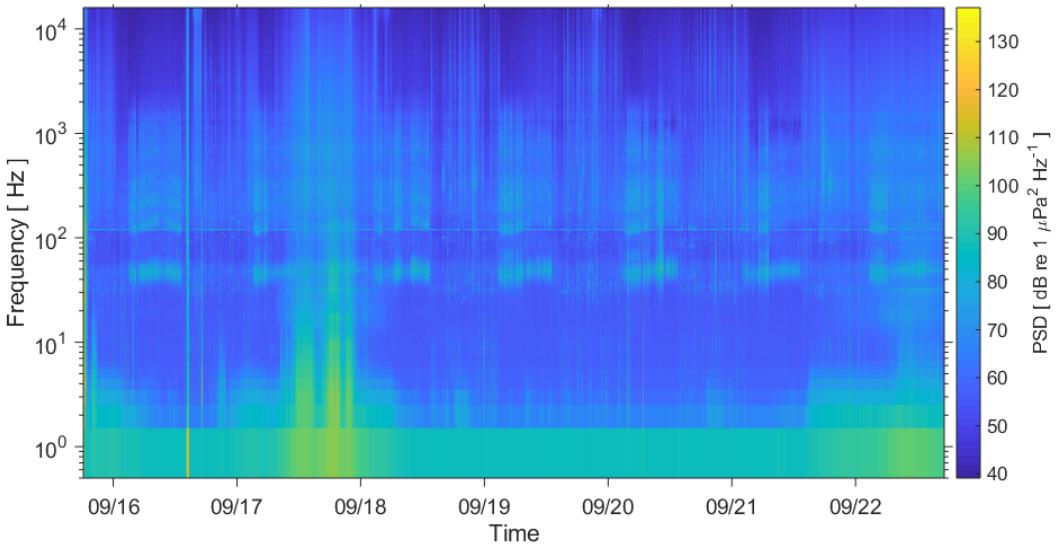


Figure 12: Long-term spectrogram of sound levels measured at Hornby Island from 15 Sep. to 22 Sep. 2019 (FFT size: 32,0000 samples, overlap: 50%, window function: Hanning).

References

- Amorim, M. C. P. (2015). Diversity of Sound Production in Fish. *Commun. Fishes*, (January 2006), 71–105. https://books.google.co.za/books?id=do0WAQAAIAAJ&hl=en&source=gbs%7B%5C_%7Dbook%7B%5C_%7Dother%7B%5C_%7Dversions
- Dosso, S. E., & Sotirin, B. J. (1999). Optimal array element localization. *J. Acoust. Soc. Am.*, 106(6), 3445–3459. <https://doi.org/10.1121/1.428198>
- Gustafsson, F. (1996). Determining the initial states in forward-backward filtering. *IEEE Trans. Signal Process.*, 44(4), 988–992. <https://doi.org/10.1109/78.492552>
- Kasumyan, A. O. (2008). Sounds and sound production in fishes. *J. Ichthyol.*, 48(11), 981–1030. <https://doi.org/10.1134/S0032945208110039>
- Kasumyan, A. O. (2009). Acoustic signaling in fish. *J. Ichthyol.*, 49(11), 963–1020. <https://doi.org/10.1134/S0032945209110010>
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science (80-)*, 220(4598), 671–680.
- Mouy, X., Rountree, R., Juanes, F., & Dosso, S. (2018). Cataloging fish sounds in the wild using combined acoustic and video recordings. *J. Acoust. Soc. Am.*, 143(5), EL333–EL333. <https://doi.org/10.1121/1.5037359>
- Mouy, X., Black, M., Cox, K., Qualley, J., Mireault, C., Dosso, S., & Juanes, F. (2020). FishCam: A low-cost open source autonomous camera for aquatic research. *HardwareX*, 8, e00110. <https://doi.org/10.1016/j.hwx.2020.e00110>
- Oppenheim, A. V., & Schafer, R. W. (2009). *Discrete Time Signal Processing 3rd Edition*. Pearson.
- Suzuki, S., & Be, K. A. (1985). Topological structural analysis of digitized binary images by border following. *Comput. Vision, Graph. Image Process.*, 30(1), 32–46. [https://doi.org/10.1016/0734-189X\(85\)90016-7](https://doi.org/10.1016/0734-189X(85)90016-7)