UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL

A two-array system for localization of fish using passive acoustic data

André Onofre Baptista Lourenço de Matos

Mestrado em Bioestatística

Dissertação orientada por:
Prof. Doutor Paulo Jorge Quintais Cancela da Fonseca
Prof. Doutor Tiago André Lamas Oliveira Marques

2021

This research was funded by Science and Technology Foundation, Portugal through the project FISHNOISE - PTDC/BIA-BMA/29662/2017

With the institutional and logistical support of
cE3c - Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa
and
MARE – Marine and Environmental Sciences Centre, ISPA-Instituto Universitário, Lisbon, Portugal.

Acknowledgments

I would like to acknowledge and thank all the people who officially and unofficially guided and helped me with the execution and writing of this work, first and foremost Prof. Paulo Fonseca, Prof. Tiago Marques and Prof. Clara Amorim, and also Manuel Vieira and Marta Bolgan. Thank you for your patience, your willingness and perseverance in sharing your knowledge and advice (even when I failed to understand and assimilate it), your time and help with the field work, all of which were essential to have what is presented in this text. I must also thank Thibault Marin-Cudraz, who helped me with valuable advice too, although he had less of an opportunity to do so, given his recent arrival to the lab. I would also like to thank Prof. Marília Antunes both for the advice and the books lent, which were very useful in helping me organise my ideas.

A special word of gratitude goes to Prof. Paulo Fonseca and Prof. Clara Amorim, for giving me the opportunity to work in their *FishNoise* project. Being paid by the project allowed me to dedicate exclusively to this work, giving me the opportunity to learn much more than I could ever hope for otherwise. I truly hope that this work's contribution to the project may live up to your trust and investment.

On a more personal note, I would like to thank my family, especially my mother, for supporting my education through all its turns and setbacks, and my girlfriend who shared with me a home, a workspace, and a lot of love and patience during this challenging time.

Finally, I would like to thank Portuguese Air Force Base number 6 for allowing the field work to collect the data analysed in this study.

Resumo

Os ambientes aquáticos costeiros são habitualmente caracterizados por baixa visibilidade, ao passo que no oceano aberto (com uma profundidade média de 3.7 km) não penetra qualquer luz para além das primeiras centenas de metros. A água tem maior impedância e menor elasticidade que o ar, o que resulta numa propagação mais rápida do som (que depende da profundidade da água e da frequência do som). Consequentemente, muitos animais aquáticos evoluíram de forma a explorar os sons como meio de percecionar o ambiente que os rodeia, encontrar comida e território, escapar de predadores e comunicar com conspecíficos. Concretamente, está demonstrado que mais de 800 espécies de peixes produzem sons, sendo que a comunicação vocal está frequentemente associada a interações sociais tais como disputas territoriais, competição por alimento e reprodução.

A biofonia (conjunto dos sons produzidos por seres vivos num determinado ambiente) no estuário do Tejo é fortemente dominada pelas vocalizações de duas espécies de peixe: a Corvina (*Argyrosomus regius,* Asso 1801) e o Xarroco (*Halobatrachus didactylus,* Bloch & Schneider 1801). *Argyrosomus regius* é um membro da família *Sciaenidae*, uma das maiores famílias de peixes capazes de produzir sons como forma de comunicação. Na época da reprodução, esta espécie forma agregados na coluna de água enquanto emite vocalizações. Pensa-se que estas tenham a função de anunciar a formação das agregações e o início da corte. *H. Didactylus* pertence à família *Batrachoididae*. Esta espécie tem um repertório de sons surpreendentemente variado e, na época de reprodução, o macho coloniza e defende ninhos em águas estuarinas pouco profundas. A partir desses ninhos, emite vocalizações denominadas sirenes (*boatwhistles)* de modo a anunciar a sua presença e atrair fêmeas para o ninho. As fêmeas depositam os seus ovos no teto do ninho, os quais são posteriormente fecundados e protegidos pelo macho, até que os alevins adquiram natação independente e dispersem.

O desenvolvimento e aperfeiçoamento de estratégias que permitam localizar peixes tirando partido dos sons que produzem pode ajudar a responder a importantes questões, inclusivamente relevantes para a sua conservação. Compreender a posição e movimentos dos cardumes durante a reprodução, assim como a forma como reagem à passagem de embarcações, por exemplo, pode ser uma importante contribuição para o delineamento de estratégias que diminuam o impacto do ruído antropogénico nestes animais. Por outro lado, saber a localização de um animal quando produz uma vocalização detetada por um hidrofone ou conjunto de hidrofones permite estimar a intensidade do som produzido. O conhecimento da intensidade sonora das vocalizações produzidas pelos animais de uma determinada espécie não só é essencial para avaliar o possível impacto do ruído antropogénico na comunicação entre conspecíficos como pode facilitar a estimação da densidade de animais a partir de monitorização acústica passiva. A *Monitorização Acústica Passiva* (*PAM*) consiste na obtenção de informação acerca das populações e ecossistemas através da análise dos sons presentes num determinado local, registados por hidrofones. Uma das grandes dificuldades que impede uma utilização mais generalizada da PAM, nomeadamente na monitorização das populações de peixes vocais é a disponibilidade muito limitada de bibliotecas de referência que permitam associar os sons registados às espécies dos animais que os produzem. Também neste aspeto a localização acústica passiva pode ter um papel importante, na medida em que é possível associá-la a sistemas de vídeo de forma a identificar que animais produzem que vocalizações e assim construir as referidas bibliotecas. Finalmente, em certas circunstâncias, é razoável assumir que os animais permanecem estáticos durante um determinado período de tempo. Tal é o caso, por exemplo, dos machos de *H. didactylus* enquanto vocalizam para atrair fêmeas. Nesses casos, saber a posição em que uma vocalização é produzida corresponde efetivamente a

saber que indivíduo a produziu. Com essa informação, é possível estudar comportamentos relacionados com a competição reprodutiva entre machos, tais como se existe sincronização ou alternância de vocalizações entre machos próximos.

As técnicas de localização por acústica passiva utilizam o atraso de um som detetado em múltiplos recetores (hidrofones), organizados num ou múltiplos *arrays,* para estimar a posição, distância ou direção da fonte sonora. Estas técnicas têm aplicação recorrente no estudo da distribuição e comportamento de mamíferos marinhos, e têm ganho cada vez mais expressão (embora esta seja ainda bastante modesta) no estudo do comportamento de peixes e na caracterização dos sons por eles produzidos.

Diferentes constrangimentos relacionados com a questão específica a que se procura responder, o local do estudo, as características do som de interesse, o equipamento disponível, entre outros, ditam uma série de escolhas que fazem com que cada implementação destes métodos tenha as suas características próprias. Tais escolhas incluem a geometria e extensão do *array* de hidrofones, a forma como é feita a medição do atraso entre hidrofones, o modelo geométrico para a propagação do som, a forma como se modela a incerteza e a estrutura de dependência entre as variáveis, e o algoritmo usado para ajustar o modelo aos dados.

Com este trabalho, foi possível desenvolver e testar um sistema capaz de estimar a posição de peixes em duas dimensões (ignorando a profundidade), que informa de modo claro e intuitivo acerca da incerteza associada a essa estimação. Este sistema é constituído por duas estruturas cúbicas de aço com cerca de um metro de aresta, por oito hidrofones, quatro dos quais foram fixos a cada uma das estruturas, pelo equipamento que regista o *output* dos hidrofones, por um altifalante submersível e amplificador associado (altifalante de calibração), e pela *pipeline* de análise utilizada para tratar os dados registados. Os dados apresentados neste estudo foram obtidos numa sessão de teste realizada numa praia dentro da Base aérea nº6 da Força Aérea Portuguesa, no Montijo (margem Sul do estuário do Rio Tejo). Aí, foi possível posicionar manualmente no fundo quer os *arrays* quer o altifalante de calibração, numa zona onde se sabia que vários xarrocos tinham estabelecido ninhos. Primeiro procedeu-se à gravação do *output* dos hidrofones enquanto era realizada a reprodução em *loop*, a partir do *altifalante de calibração* de um *boatwhistle* de xarroco préviamentepreviamente gravado; de seguida gravou-se a *paisagem sonora* que incluía vocalizações de vários xarrocos presentes na área circundante. Os dados obtidos da reprodução foram utilizados para corrigir erros de posicionamento dos *arrays* e quantificar a incerteza associada a esse posicionamento, e os dados provenientes da paisagem sonora natural foram utilizados para localizar os xarrocos que produziram vocalizações durante o período de gravação. A análise subsequente permitiu obter regiões com credibilidade superior a um valor predefinido para a posição dos animais. Essas representam uma região do plano horizontal sobre (ou sob) a qual o animal se encontrava no momento da vocalização com probabilidade superior a um valor arbitrário (no caso dos resultados apresentados $p \geq 0.961$).

Os resultados obtidos são promissores no sentido em que são consistentes com o que se esperava tendo em conta o conhecimento prévio do local de teste. Em particular, o facto de ter sido possível discriminar a partir dos dados de cada *array,* grupos de sons provenientes de diferentes direções, e esses grupos serem consistentes entre *arrays* (apesar de os dados terem sido obtidos independentemente por cada *array*) é consistente os machos de xarroco vocalizarem repetidamente a partir de posições fixas (os seus ninhos).

Uma vez que os estudos que utilizam acústica passiva para localizar peixes são ainda escassos, e que os métodos utilizados são fortemente dependentes das características particulares de cada estudo, é difícil comparar diretamente o método apresentado aqui com outros métodos já publicados. No entanto, merecem nota algumas particularidades da solução apresentada, nomeadamente: (1) a utilização de múltiplos *arrays*

para ultrapassar a limitação da localização pontual a animais que se encontrem a distâncias do *array* da mesma ordem das distâncias entre hidrofones; (2) a produção de um *output* facilmente interpretável quer no que toca à localização do indivíduo que vocaliza, quer no que toca à incerteza associada; (3) a modelação da estrutura de dependência entre *lags* observados pelos diferentes pares de hidrofones que não assume que estes sejam independentes; e (4) a utilização do conhecimento do facto de determinados peixes vocalizarem a partir de posições fixas para estimar o erro máximo associado às estimações.

Esta dissertação foi realizada no contexto de um projeto mais longo, que procura compreender e descrever os efeitos do ruído antropogénico no comportamento e sucesso reprodutor da Corvina e do Xarroco no estuário do Tejo. O trabalho apresentado neste documento representa, portanto, um passo intermédio mas fundamental na obtenção de dados biológicos relevantes para essa compreensão. Apesar de algumas das limitações identificadas estarem ainda por resolver, acreditamos que será brevemente possível usar este sistema para aprofundar o nosso conhecimento acerca das questões que motivaram o seu desenvolvimento.

Palavras-chave: acústica passiva; localização; *Argyrossomus regius; Hallobatrachus didactylus*; Estuário do Tejo.

Abstract

Passive acoustic localization methods use the time lag of a sound detected at several receivers (or at a single receiver via directed and reflected paths) to estimate the position, range or direction of a sound source. Even though the use of this method applied to vocal fish is still scarce in scientific literature, it has important potential applications, such as the elucidation of courtship and spawning behaviour of vocal fish, the estimation of the source level of vocalisations, the construction of reference libraries that associate sample vocalisations to the species that produce them, and the understanding of the role of vocal communication in social interactions such as those involved in reproductive competition.

In the present text we present a low-cost system that can both estimate the positions of vocalising fish and intuitively convey the uncertainty of such estimates to the user. All the data presented in this work was obtained in a single test deployment of the recording equipment, and this dissertation was done within the context of a project with a longer temporal span. Therefore, the results presented here represent a fundamental but intermediate step towards the obtention and analysis of relevant biological data.

Since studies using passive acoustic localization for fish triangulation are still scarce, and the methods employed are constrained by their specific features, it is hard to directly compare our method to the ones existing. Yet, some features in which it differs from previous work regarding fish localization by passive acoustic methods include (1) the usage of multiple arrays to overcome the limitation of point localization to distances of the same order of magnitude as the distance between the hydrophones of the array, (2) the output of a readily interpretable result that informs on both the estimated location and the associated uncertainty, (3) the modelling of the dependency structures between observed *TDOA*s of different pairs of hydrophones, and (4) the estimation of *observed TOA* variance using fish known to vocalise from fixed positions.

Keywords: passive acoustics, localization, *Argyrossomus regius, Halobatrachus didactylus*, Tagus estuary.

Contents

Figures

Symbols

$\boldsymbol{f}$	Frequency.
$\boldsymbol{C}$	Speed of sound.
$\boldsymbol{t_c}$	Point in time in which a certain waveform peak is observed on channel *c*. ($c = 0,1,2,3$).
$\boldsymbol{x_c}'$	Time difference between a waveform peak on channel c and the same peak on channel 0 (reference channel) ($c = 1,2,3$).
$\boldsymbol{\alpha}$	Azimuth relative to the array reference system (see Figure 2.5).
$\boldsymbol{\theta}$	Elevation (see Figure 2.5).
$(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$	Cartesian coordinates relative to the *array reference system*
$(\boldsymbol{x_w}, \boldsymbol{y_w}, \boldsymbol{z_w})$	Cartesian coordinates relative to the *world reference system*
$\mathbf{HP_i}$	One of four hydrophones numbered $i = 0$ to $i = 3$.
$\boldsymbol{X}$	A random vector of 3 *TDOA* observations of a single sound, one for each of three hydrophones (HP_1, HP_2, HP_3,) relative to a fourth (HP_0).
$\boldsymbol{x}$	A realization of $\boldsymbol{X}$
$\boldsymbol{\mu}$	Mean of the random vector $\boldsymbol{X}$.
$\boldsymbol{\Sigma}$	Covariance matrix of the random vector $\boldsymbol{X}$.
$\boldsymbol{d_i}$	The distance travelled by the wave front from the moment it crosses a *non-reference hydrophone* (HP_i) to the moment it crosses the *reference hydrophone* (HP_0).
$\boldsymbol{u}$	A unitary vector representing the direction of sound propagation.
$\boldsymbol{b}$	A unitary vector, representing the bearing of the source relative to the array.
$\boldsymbol{r_i}$	The position of HP_i relative to HP_0.
$\boldsymbol{Y}$	A random vector of 4 time of arrival (TOA) observations of a single sound to each hydrophone HP_i.
$\boldsymbol{Y_i}$	TOA observation of a sound to hydrophone HP_i (a random variable).
$\boldsymbol{\sigma_Y^2}$	The variance of Y_i.
$\boldsymbol{\sigma_{prop}^2}$	The variance of Y_i due to propagation uncertainty.
$\boldsymbol{\sigma_{meas}^2}$	The variance of Y_i due to measurement uncertainty.
$\boldsymbol{\sigma_{prop+meas}^2}$	The variance of Y_i due to propagation and measurement uncertainty.
$\boldsymbol{\sigma_{hp}^2}$	The variance of Y_i due to hydrophone position uncertainty.
$\boldsymbol{\tilde{\sigma}_Y^2}$	Worst case guess for σ_Y^2.
$\boldsymbol{\tilde{\sigma}_{hp}^2}$	Worst case guess for σ_{hp}^2.
$\boldsymbol{\tilde{\sigma}_{prop+meas}^2}$	Worst case guess for $\sigma_{prop+meas}^2$.
$\boldsymbol{\tilde{\sigma}_{pos}}$	Worst case guess for the standard deviation of the position of each hydrophone relative to its planned position.
$\boldsymbol{r}$	(in the context of the E-M algorithm) Number of groups of vocalisations (putative individual males/nests).
$\boldsymbol{k}$	Specific group of vocalisations (putative individual male /nest).
$\boldsymbol{b_k}$	3D bearing of putative male *k* relative to an *array.*
$\boldsymbol{x_i}$	The i^{th} realization of $\boldsymbol{X}$.
$\boldsymbol{\mu_k}$	Mean vector of *TDOA*s for observations belonging to group k.
$\boldsymbol{\sigma_{Y_k}^2}$	Variance of the *TOA*s of observations belonging to group *k.*

$\boldsymbol{z_{ki}}$	Membership of each observation *i* to group *k*.
$\boldsymbol{z_{ki}^{(j)}}$	Estimate of z_{ki} at the j^{th} iteration of the E-M algorithm.
$\boldsymbol{p_k^{(j)}}$	Estimate of the proportion of observations belonging to group k at the j^{th} iteration of the E-M algorithm.
$\boldsymbol{p^{(j)}}$	A vector of r elements $\boldsymbol{p_k^{(j-1)}}$
$\boldsymbol{\mu_k^{(j)}}$	Estimate for the mean vector of group k at the j^{th} iteration of the E-M algorithm.
$\boldsymbol{M^{(j)}}$	A $(3 \times k)$ matrix, each column the estimate for the mean vector of group k at the j^{th} iteration of the E-M algorithm ($\boldsymbol{\mu_k^{(j)}}$).
$\boldsymbol{\sigma_{Y_k}^{2\,(j)}}$	Estimate of $\sigma_{Y_k}^2$ at the j^{th} iteration of the E-M algorithm.
$\boldsymbol{\hat{\mu}_k}$	Estimate for the mean vector of group k at the end of the E-M algorithm.
$\boldsymbol{\hat{\sigma}_{Y_k}^2}$	Estimate of $\sigma_{Y_k}^2$ at the end of the E-M algorithm.
$\boldsymbol{\hat{z}_k}$	Estimate of z_{ki} at the end of the E-M algorithm.
$\boldsymbol{r, s}$	(in the context of the Bayesian grid) number of azimuth and elevation values in the grid.
$\boldsymbol{(\alpha_i, \theta_j)}$	The centre of each cell in the grid of possible bearings.
$\boldsymbol{\breve{\sigma}_{Y_k}^2}$	Maximum likelihood estimator of $\sigma_{Y_k}^2$ for speaker data.
$\boldsymbol{\delta_\alpha}$	Difference in azimuth between consecutive grid cells.
$\boldsymbol{\delta_\theta}$	Difference in elevation between consecutive grid cells.
$\boldsymbol{s_c(\alpha_i)}$	Cumulative sum of posterior probability of azimuth values (ordered by decreasing posterior probability).
$\boldsymbol{\alpha_{low}, \alpha_{up}}$	Limits of the HPD credible interval for the azimuth of a sound source relative to an array.
$\boldsymbol{\alpha'_{sp}}$	Azimuth of the speaker relative to an array, were the array aligned to the *world reference system*.
$\boldsymbol{\alpha_{sp}}$	Azimuth of the speaker relative to the *array reference system.*
rot	The rotation needed to bring the array into alignment with the *world reference system*.
$\boldsymbol{\hat{\alpha}}$	An estimate of the azimuth relative to the *array reference system*.
$\boldsymbol{\alpha'}$	The azimuth of a sound source relative to an array, were it aligned to the *world reference system.*
$\boldsymbol{\hat{\alpha}'}$	An estimate of α'.
$\boldsymbol{\alpha_{sp_{low}}, \alpha_{sp_{up}}}$	Limits of the credible intervals for the azimuthof the speaker (α_{sp}).
$\boldsymbol{rot_{low}, rot_{up}}$	Limits of the credible intervals for rot.
$\boldsymbol{A}$	Event: α is in the interval $[\alpha_{low}, \alpha_{up}]$.
$\boldsymbol{R}$	Event: rot is in the interval $[rot_{low}, rot_{up}]$.
$\boldsymbol{A'}$	Event: α' is in the interval $[\alpha'_{low}, \alpha'_{up}]$.
$\boldsymbol{p_a}$	Probability associated to the credible interval $[\alpha_{low}, \alpha_{up}]$.
$\boldsymbol{p_r}$	Probability associated to the credible interval $[rot_{low}, rot_{up}]$.
$\boldsymbol{\alpha'_{low}, \alpha'_{up}}$	*Lower-bounded credible interval* with minimum probability $p_s = p_\alpha p_r$.
$\boldsymbol{p_s}$	Lower bound for the probability associated the *lower-bounded credible interval* $[\alpha'_{low}, \alpha'_{up}]$
$\boldsymbol{a_1, a_2}$	Areas on the horizontal plane between the 2 rays originating at each array's position with directions defined by the corresponding limits: α'_{low} and α'_{up}.

$\boldsymbol{a_c}$	Intersection of a_1 with a_2
$\boldsymbol{C}$	Event: the vertical projection of the position of the sound source onto the $x_w 0 y_w$ plane is included in the area a_c;
$\boldsymbol{A_1'}$	Event: the vertical projection of the sound source onto the $x_w 0 y_w$ plane is included in the area between the 2 rays originating at $(x_{a1}, y_{a1}, 0)$, with directions defined by α'_{low_1} and α'_{up_1};
$\boldsymbol{A_2'}$	Event: The vertical projection of the sound source onto the $x_w 0 y_w$ plane is included in the area between the 2 rays originating at $(x_{a2}, y_{a2}, 0)$, with directions defined by α'_{low_2} and α'_{up_2}.
A, B and C	Putative individual fish, assumed to be responsible for each group of sounds discriminated by the E-M algorithm. Also the groups themselves. *Roman text* is used to avoid ambiguity with A and C, which refer to events, and are denoted by *italic text*.

Abbreviations

TOA	Time of arrival
TDOA	Time difference of arrival
PAM	Passive acoustic monitoring
SNR	Signal-to-noise ratio
SPL	Sound pressure level
E-M	Expectation-Maximization (algorithm)
psu	Practical salinity units
HPD	Highest posterior density

1. Introduction

Aquatic environments are often characterised by poor visibility in coastal areas, while the open ocean, whose average depth is 3.7 km, is pitch black below a few hundred meters. Water has higher impedance and lower elasticity than air, which generally results in faster and better propagation of sounds in underwater environments. As a result, many aquatic animals have evolved to use sounds for sensing their environment, finding food and territories, escaping from predators and for communicating with conspecifics (Hawkins and Popper, 2018; Slabbekoorn et al., 2010). These sounds are nowadays routinely used to make inferences about wildlife. In particular, more than 800 species of fish have been documented to produce sounds, with vocal communication being often associated to social interactions such as territorial disputes, competition for food, courtship, and spawning (Radford et al., 2014). The sounds produced by animals in a given environment constitute its *biophony*. Together with those, the sounds produced by geophysical phenomena (*geophony*) and those produced by human activity (*antropophony*) make up the *soundscape* of the environment (Pijanowski et al., 2011).

1.1. What is sound?

Sound results from vibrating structures that cause oscillation of the medium particles and alternating regions of high and low pressure that propagate away from the source (sound waves). The oscillation of the medium particles may be detected by fish and constitutes the *particle motion* component of the sound wave. The alternating regions of high and low pressure make up the *pressure* component of the sound wave which is detected by fish with morphological adaptations connecting the swimbladder to the inner ear endorgans, and is the sound component registered by hydrophones. The input of the hydrophone is the oscillating pressure at its position, and its output is an oscillating voltage directly proportional to that pressure. (Au and Hastings, 2008)

A representation of the pressure detected (or equivalently apart from scale, the voltage produced) by a hydrophone over time constitutes the *waveform representation* of the sound (Brüel&Kjær, 2019). If such a representation is a sinusoidal wave, the sound is said to be a pure tone. Such a tone is unequivocally defined by a *frequency*, an *amplitude* and a *phase*. *Frequency* is the number of oscillations per unit time, *amplitude* is the maximum deviation of the pressure from the equilibrium pressure, and *phase* defines a temporal offset. Any periodic sound which is not a pure tone may be described as the sum of a number of pure tones of given frequencies, amplitudes and phases. Those theoretical sinusoidal tones are referred to as the frequency components of a sound. A sound with frequency components that include a wide range of frequencies is said to be broad-band, as opposed to narrow-band sounds which have few components with relatively similar frequencies (the extreme case of a narrow-band sound is a pure tone). Sounds are often composed of several components whose frequencies are multiples of the lowest one. In such cases the lowest frequency is called *fundamental frequency* and the successive multiples are called *harmonics*. The *fundamental frequency* may or may not be the *dominant frequency* which is the component with the highest intensity (or amplitude) (Randall, 1987). The scaling constant that relates the magnitude of the input to that of the output of a recording system is its *gain* (Brüel&Kjær, 2019).

Sounds are often distorted by broad-band ambient noise and/or narrow-band electrical noise. The *Signal to Noise Ratio* (SNR) is the ratio between the intensity of the signal relative to that of the noise. In the context of this work, the term *signal* refers to the sound emitted by the vocalising fish. By removing from the

recording the frequency components that are not represented in the signal and/or are over represented in the noise, the SNR increases (Randall, 1987). Clearly this is desirable since the noise carries no information and generally increases the variance of estimates obtained from the recording. Selectively removing or decreasing the energy of certain frequency bands is called *filtering* and is a usual step in pre-processing sound files for source localization (Randall, 1987).

The intensity of a sound is its energy per unit area and is directly proportional to the square of the pressure (amplitude). Furthermore it is usually represented on a relative logarithmic scale and referred to as the *sound pressure level (SPL)* (Brüel&Kjær, 2019). A representation of the intensity of the different frequency components over time is called a *spectrogram.* In such representations, the time corresponds to the *x* axis, the frequency to the *y* axis and the intensity (or, alternatively the amplitude) of the sound is represented by colour (Randall, 1987).

1.2.Propagation of sound in water

Sound waves travel away from the sound source at a speed that depends on the physicochemical characteristics of the medium (Au and Hastings, 2008). In the case of sea or estuarine water, the relevant parameters are the hydrostatic pressure, temperature and salinity.

Considering a sound propagating from a point source in all directions at a given speed, and assuming the water is not displaced by fast currents, we may visualise the wave front as a spherical surface. Yet, when the sound is received far away from the source, only a small fraction of the surface is perceived (see Figure 1.1). In these conditions, the portion of the wavefront detected (e.g. by an array of hydrophones) is approximately planar.

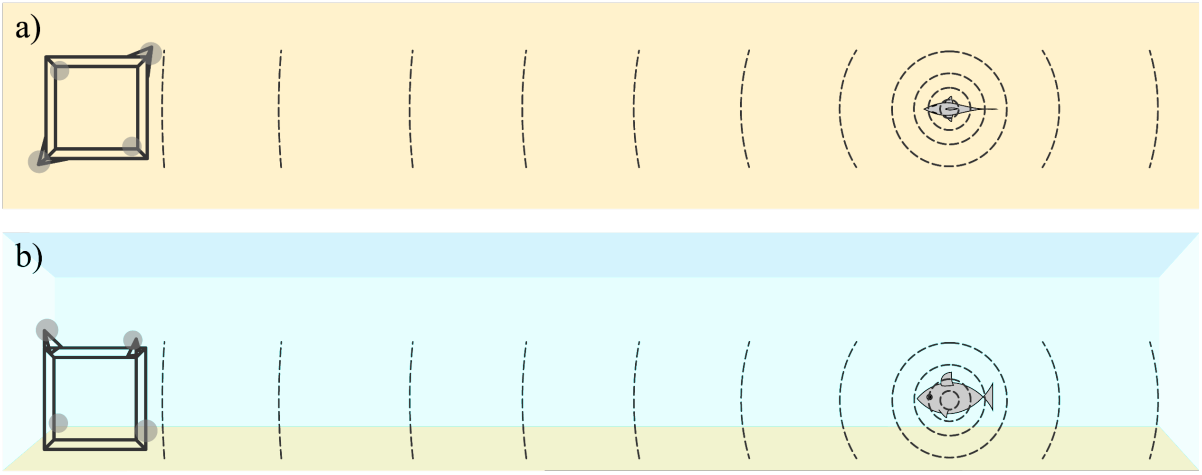

Figure 1.1 ***S*chematic representation of the wave front propagating (directly) from the vocalising fish to the array**. a) As seen from the top; b) as seen from the side. When the array of hydrophones is far from the source of the sound, the portion of the wave front that is perceived by the array is approximately planar.

When the wave front is modelled as a spheric surface, if the position of two hydrophones is known, the time difference of arrival (*TDOA*) of a sound to one hydrophone relative to another, gives a hyperboloid as the region where the source lies. With three such pairs (forming an array), it is possible to locate the source to a specific point in 3D space by finding the intersection of the 3 hyperboloids (see Figure 1.2).

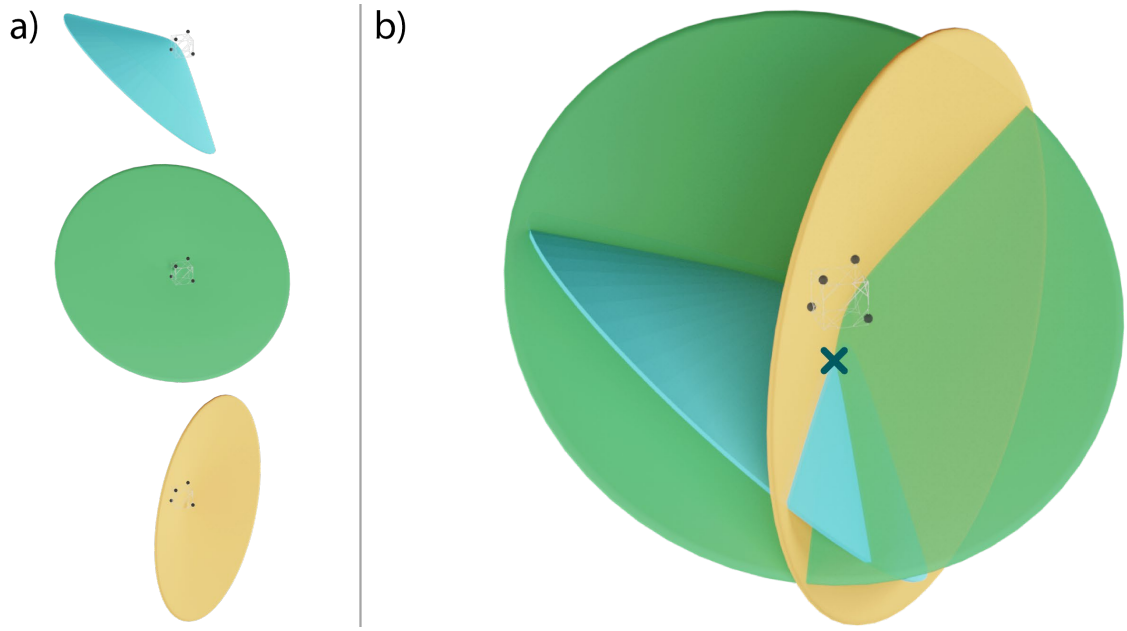

Figure 1.2 **Example hyperboloids corresponding to the possible positions of a fish modelled from the *TDOA* at each pair of hydrophones of a four-hydrophone array (assuming spherical propagation).** a) There is a TDOA relative to each pair of hydrophones. To each of those, corresponds a hyperboloid in which the fish is estimated to be. b) With three pairs of hydrophones, the intersection of the corresponding hyperboloids is a point, which is the estimated position of the fish in 3 dimensions (marked by a dark blue ***X***). The grey spheres mark the positions of the hydrophones.

On the other hand, if the wave front is modelled as a planar surface, each pair of hydrophones and corresponding *TDOA* yields a region of possible source positions that corresponds to the lateral surface of a cone with apex between the 2 hydrophones and infinite height. If the source is far from the array, the apices of such cones approximately coincide. Therefore, their intersection is not a point but a ray (see Figure 1.3). That ray represents a 3D bearing and may be defined by the position of the array and two angles: azimuth and elevation. These angles are more formally defined in the section 2.4 but for now it is sufficient to note that azimuth describes a direction on the horizontal plane, while elevation describes the vertical angle that the incident sound wave makes with that plane.

Besides the direct path from the source to the array, the sound waves may be detected by the array after being reflected off the surface and/or bottom. If the sound arrives at the hydrophones with significant energy by both the direct and reflected paths, the sound pressure registered at each hydrophone's position is the sum of the pressures due to each of the two wavefronts. This causes distortion of the waveform in a phenomenon called interference. If one of the signals is much stronger than the other, there will be no apparent distortion, but the sound will seem to come from the direction of the stronger path (be it the reflected or the direct one). On the other hand, if both have similar energy, the shape of the waveform will be different for each hydrophone and it may be hard to discern the direction of any of the paths. If the reflected path does dominate over the direct one, the elevation with which the sound wave impinges on the array will not be the true elevation of the source, but will instead be that of the point where the wave was last reflected. Note however, that if both the bottom and the surface are assumed planar and horizontal only the apparent elevation of the source will be affected by a dominant reflected path, and not its apparent azimuth (Figure 1.4).

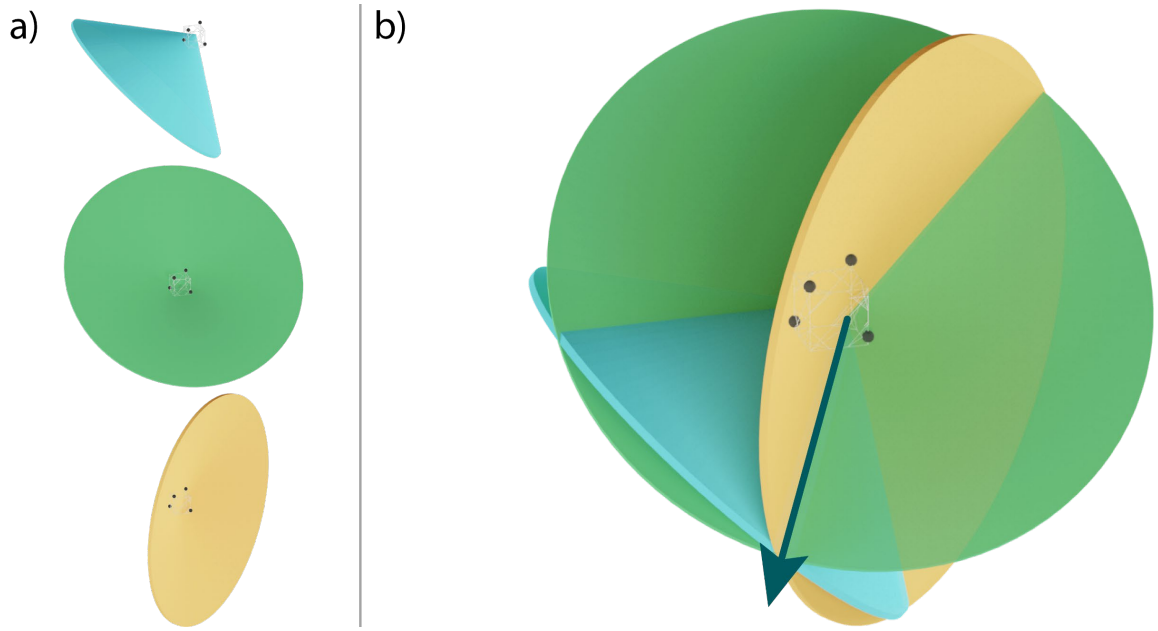

Figure 1.3 **Example conic lateral surfaces corresponding to the possible 3D bearings of a fish modelled from the *TDOA* at each pair of hydrophones of a four-hydrophone array (assuming planar propagation).** a) There is a TDOA relative to each pair of hydrophones. To each of those, corresponds a lateral conic surface which is estimated to include the direction of the ocalising fish. b) With three pairs of hydrophones, the intersection of the corresponding lateral conic surfaces ray (marked by a dark blue arrow), which is the estimated bearing of the fish relative to the array. The grey spheres mark the positions of the hydrophones.

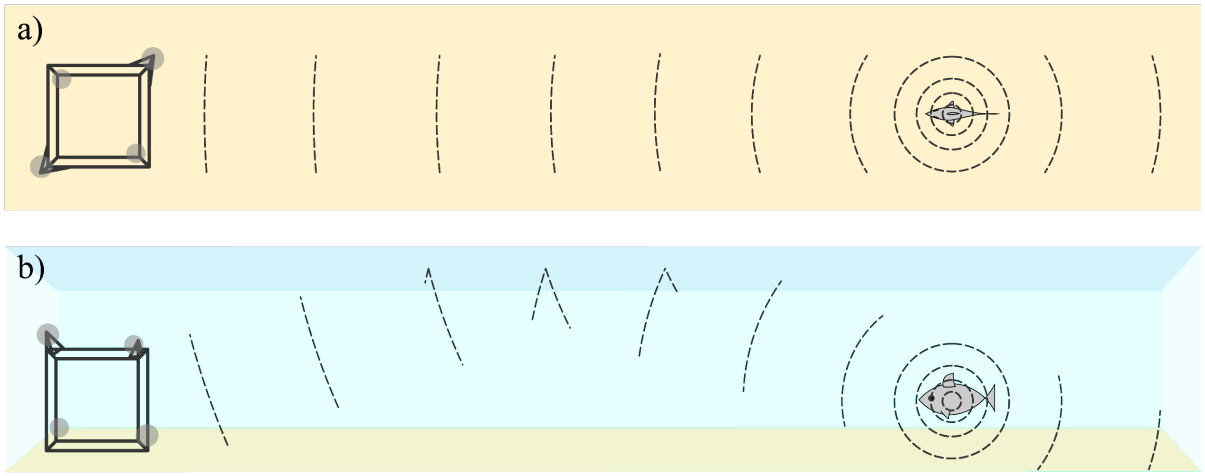

Figure 1.4 ***S*chematic representation of the wave front propagating from the vocalising fish to the array, via a reflected path**. a) As seen from the top; b) as seen from the side (b). Even though a wave that has been reflected off the surface arrives at the array from an elevation higher than the elevation of the fish relative to the array, it arrives from the same azimuth (direction over the horizontal plane) as if it had followed a direct path.

A sound received at a greater distance from its source is perceived less intensely than it would, if received nearer to the source. This is due to two phenomena: some energy is dissipated during propagation, and the remaining is spread through an increasing area, of which only a fraction affects the hydrophone. Since the intensity of a sound received at an hydrophone depends on the distance from the source to the hydrophone, it doesn't inform on how powerful a sound source is. In order to convey that information, the *source level* (by convention the *sound pressure level* (SPL) as received $1m$ away from the source) is used. Note that it

is possible to model the SPL at any location as a function of the source level and distance from the source (Au and Hastings, 2008), which may have several relevant applications, some of which are briefly discussed in 1.5.1.

1.3. Recording quality validation

Ideally the output (voltage) of a hydrophone would instantly mirror its (pressure) input, regardless of the frequency of the sound. Real hydrophones, however, respond differently to sounds with different frequencies. Particularly, both the *gain* and the *phase delay* of the response may vary according to frequency. The *gain* and the *phase delay* as functions of frequency are termed *amplitude frequency response* and *phase frequency response*. If these vary widely across the frequencies of the sound of interest (in our case toadfish and meagre vocalisations), the waveform is distorted and the *TDOA* measurements may be compromised. It is of particular importance for *TDOA*-based localization that the phase response is similar for the different hydrophones used. This is because it describes the delay with which the pressure oscillations are transduced to voltage oscillations. If this delay is different across hydrophones, a bias will be introduced in the *observed TDOA*s of each hydrophone pair.

To characterise the *phase* and *gain* frequency response of a hydrophone, one usually compares its output with that of a high quality hydrophone. The output of the high quality hydrophone is assumed to be the same as the input of the hydrophone being analysed. By computing a *Fast Fourier Transform (FFT)* of each of the outputs and computing their ratio we obtain a *transfer function* that describes both the *amplitude* and *phase frequency responses* of the hydrophone. Flat functions mean that the response of the system (output) varies similarly to the input signal in amplitude and phase across frequency. A zero value of the transfer function gain (dB) and phase (degree) means no amplitude difference and no phase shift of the output relative to input, respectively. If such a *transfer function estimate* is computed multiple times, it is possible to estimate its random variability and describe it also as a function of frequency. The function typically used for this is called *coherence* and represents how much of the output of the hydrophone is explained by its input. A value close to 1 validates the results obtained regarding the *amplitude* and *phase frequency responses*. On the other hand, if for a certain frequency the coherence is much smaller than 1, this indicates that the response of the hydrophone is not consistent across the trials, and so it is not possible to confidently assert the *frequency response* of the hydrophone at that frequency.

1.4. Active acoustic methods for fish localization

Due to the high efficiency with which sound propagates in water, many methods to monitor fish (and other aquatic animals) abundance, location and movement are based on sound propagation. These include active sonar methods, active acoustic localization tags and passive acoustic monitoring.

1.4.1. Sonar

Sonars and echosounders are devices that emit soundwaves and analyse their echoes to detect objects, including fish shoals, and characterise bottom features (Misund, 1997). Sonar methods are usually employed from research vessels to conduct population surveys, and are known to incur in bias, namely due to vessel avoidance (De Robertis and Handegard, 2013). Besides the obvious negative bias inherent to the avoidance, there are instances in which the effect is not as obvious, such as when the strength of the signal is increased due to the tilt change of fish diving (Hazen and Horne, 2004; Nakken and Olsen, 1977).

Although commercial fishing sonars operate at frequencies outside the hearing range of most fish (usually higher than 20 kHz), they often produce very high source levels, and some fish have been shown to detect sounds with such characteristics (Mann et al., 2001). This means the sound emitted by the sonar itself may also affect the behaviour of fish. Furthermore, some naval low- and midfrequency sonars use sounds from several hundred to several thousand hertz (which can have deleterious effects on fish since they generate high intensity sounds within the hearing range of many fish) (Hawkins and Popper, 2018).

1.4.2. Acoustic telemetry

Acoustic telemetry methods have been successively employed, especially in the last 20 years, to unveil and quantify previously unobserved processes important to population dynamics, reproductive performance, and fitness in a wide range of taxa, namely fish and other aquatic organisms (Crossin et al., 2017). In its simplest form, acoustic telemetry consists of a tag that is implanted or attached to an animal, and actively emits a sound pulse that is detected and logged by one or several receptors. The time of arrival of the signal at multiple synchronized detectors can be used to locate the fish. Individual tags may be coded to transmit individual IDs, and sensors may be incorporated in the tag to transmit environmental data. Besides being expensive compared to passive acoustic monitoring, these methods invariably require intrusive manipulation of the animals, for implantation of the tags, which can arguably affect their behaviour and fitness.

1.5. Passive acoustic localization

Passive acoustic localization methods use the time lag of a sound detected at several receivers (or at a single receiver via directed and reflected paths) to estimate the position, range or direction of a sound source. It has been used for some time to study marine mammal distribution and behaviour (Bénard et al., 2011; e.g. Clark et al., 1996; Giraudet and Glotin, 2006; Glotin et al., 2008; Helble et al., 2015; Nosal and Frazer, 2007; Roy et al., 2010; Wang et al., 2016), and is increasingly being used to study fish behaviour (Parsons et al., 2009, 2010; Putland et al., 2018) and characterise fish sound production (Locascio and Mann, 2011; Mouy et al., 2018).

1.5.1. Applications

Even though the use of this method applied to vocal fish is still scarce in the scientific literature, it has important potential applications for conservation. One such application is the elucidation of courtship and spawning behaviour, which may contribute to the management of exploited species (for instance by permitting proper timing of seasonal fishery closures designed to protect spawning fish (Parsons et al., 2009). By integrating fish behaviour data obtained with passive acoustic methods with data on human activities, like local boat traffic, it may also be possible to uncover potential impacts of those activities on fish reproduction and fitness. Another application is the estimation of the source level of vocalisations. If the location of the source is known (or estimated using passive acoustic localization), so is its distance to the hydrophones, and it is possible to use that information together with the received SPL to calculate the source level (Locascio and Mann, 2011). Source level is an important feature of fish vocalisations because, together with habitat-dependent signal transmission loss, background sound levels, and hearing sensitivity, it allows the estimation of the range of effective communication between soniferous fishes. Also, having the typical source level of a specific species documented may allow the use of data obtained with single hydrophones to estimate ranges from received SPL. Such range estimates may then be used to estimate population density (Marques et al., 2013).

Many animals emit acoustic signals that encode information about their presence and activity. This offers an opportunity to use Passive Acoustic Monitoring (PAM) to acoustically survey wildlife and environment. PAM has experienced an important growth as the associated technological costs and constraints decreased in recent years (Gibb et al., 2019). One of the greatest constraints to a broad use of PAM, namely for monitoring fish, is the limited availability of reference sound libraries (needed to associate fish species to recorded sounds). This is partly due to behaviour-related sounds produced in natural habitats being often difficult or impossible to induce in captivity. On the other hand, *in situ* measurements are challenging and require the accurate localization of the fish, since there is no control over such variables as the number of species present or the number of fish vocalising (Mouy et al., 2018). Therefore, passive acoustic localization may have an important role bridging that gap.

Finally, in cases where the vocalising fish may be considered stationary (such as the male toadfish after occupying a nest (Amorim et al., 2016), a location may be assigned to an individual, which means some questions about the behaviour associated with reproductive competition between males may be easily addressed. Such questions include whether males interact vocally with each other (e.g. by alternating or synchronising calls, or when using sounds to defend their nests).

1.5.2. Implementation

Implementations of passive acoustic localization methods differ in the geometry and extension of the array of hydrophones, in the way the time lag is measured (e.g. onset detection, peak detection, or cross-correlation), in the way the geometry and uncertainty of the problem are modelled, and in the algorithm used to fit the model to the data. Different considerations may influence the choices for each particular implementation, including the characteristics of the sounds produced by the object of the study, the available equipment, computational efficiency, possibility of automation, ease of interpretation of the outputs, and plausibility and validity of required assumptions.

Observed lag measurement may follow two broad approaches. Either the time of arrival (*TOA*) of the sound at each hydrophone is measured and the *observed TDOA*s are computed as the differences between the *observed TOA*s, or the *observed TDOA*s are measured directly.

In (Gervaise et al., 2019), the authors define the time of arrival as the moment at which the energy of the received signal crosses a threshold. This approach lends itself very well to automation, but depends on the sounds having a steep, clear onset. Still under the *TOA* approach, both Putland et al. (2018) and Locascio and Mann (2011) use a different strategy, by defining the *observed TOA* as the first amplitude peak of the signal. The implementation of this strategy may (Locascio and Mann, 2011) or may not (Putland et al., 2018) be automated, but is again limited by the sound having an onset that is clear enough for this first peak to emerge from background noise.

Mouy et al. (2018) on the other hand, measured the *TDOA*s by computing the *cross correlation function* of the sounds recorded by each pair of hydrophones. The *cross correlation function* is obtained by computing the correlation between the output of one hydrophone and the output of a second hydrophone as a function of a *lag* applied to one of the outputs. The *lag* that maximises the *cross correlation function* is then taken as the estimate for the *TDOA*. This approach uses information conveyed by the whole length of the sound instead of just the first peak, and so it does not require the sound to have a steep onset, but it has two drawbacks. The first is that, if the sound is reflected on nearby surfaces, the reflections may be detected by the hydrophones and the latter part of the output is distorted (a phenomenon known as reverberation). The

second is that for long, relatively narrow-band sounds such as the vocalisations of individual Meagre and of Lusitanian toadfish, the spacing between hydrophones is limited to $C/2f$ (where C is the speed of sound and f is the fundamental frequency). If the distance is greater than $C/2f$ there will be more than one local maximum, and the estimate of the *TDOA* might be ambiguous.

Generally, the way the geometry is modelled depends on the spacing between hydrophones relative to the distance of the sound source to the hydrophones. When they are similar, the wave front is modelled as a spherical surface, and an estimation of the 3D position of the source is possible. In cases where the source is much further from the hydrophones than the hydrophones are to each other, the wave front is approximated by a plane, and only the bearing can be accurately estimated. For logistical reasons, the implementation presented in this work was limited to compact arrays of hydrophones (*circa* $1.4\ m$ distance between hydrophones), while the targets of the study (*Argyrosomus regius* and *Halobatrachus didactylus*) are fish that can be heard at least up to distances of tens of meters. This meant that the appropriate geometrical model in our case was the planar wave model. Even though the planar wave model only allows for bearing estimation, an alternative strategy was devised to enable obtaining an accurate position in 3D. This involved the deployment of two arrays, the obtention of credible regions for the bearings of the fish relative to each array, and the computation of the intersection of those regions, taking into account the uncertainty in array positioning. The result is a *region* that includes the position of the vocalising fish with a probability *higher than* an arbitrary pre-defined value. Such a region is different from a *credible region* in that the latter would have a defined probability of containing the true position of the fish, instead of a defined *lower bound* for that probability. Therefore, for the remainder of this text, we will refer to these regions as *lower-bounded credible regions.*

To the best of our knowledge, no published work on fish localization uses credible regions of single sounds to describe the uncertainty associated to the position estimates. Instead, the uncertainty is described for the whole experiment, usually assuming the *observed TDOA*s for each pair of hydrophones are independent and the standard error of the *TDOA estimates* is constant (Gervaise et al., 2019; e.g. Locascio and Mann, 2011; Mouy et al., 2018). Nosal and Frazer (2007) use *TDOAs* to track Sperm whales and produce an approximated *conditional confidence interval* for each coordinate of the animal's position. To do this, the coordinates are discretised, and the likelihood at each point of the grid is computed in order to obtain an approximate maximum likelihood estimate. This likelihood is maximised over the 3D space, to obtain a source location estimate, but not over the possible values of *observed TOA* variance. Instead, it is made conditional on a *"worst case" standard deviation* based on *best guesses* for the different sources of uncertainty (namely *TOA* measurement error, receiver position uncertainty, and sound propagation variability).

1.6. Vocal fish in the Tagus estuary

The Tagus estuary biophony is dominated by the advertisement calls of two fish species, the Meagre (*Argyrosomus regius,* Asso 1801) and the Lusitanian toadfish (*Halobatrachus didactylus,* Bloch & Schneider 1801). *Argyrosomus regius* is a member of the *Scianidae* family, one of the largest families of soniferous fishes. This species forms spawning aggregations during which it produces advertisement calls. These have been suggested to serve the purposes of announcing the formation of the aggregations, and the start of courtship behaviour (Lagardère and Mariani, 2006; Bolgan et al., 2020). *Halobatrachus didactylus* is a batrachoidid with an unusually varied sound repertoire (Amorim et al., 2008). Males of this species colonise and defend nests in estuarine shallow waters during the mating season, emitting conspicuous advertisement calls (referred to as *boatwhistles*). These vocalisations are essential to attract females, who

deposit their eggs on the nest roof (Amorim et al., 2016). The male then fertilises the eggs and provides parental care until the fry is free-swimming (Dos Santos et al., 2000). As mentioned in 1.5.1, devising strategies to locate vocalizing fish in the natural habitat can allow to address important questions such as the position of the aggregates relative to the substrate and nearby structures, the effect of passing vessels on the behaviour of the animals, and how individuals interact vocally.

1.7.Aim

Our goal was to build and test a low-cost system which could both estimate the positions of vocalising fish and intuitively convey the uncertainty of such estimates to the user. In order to achieve this goal, we projected a system that included the recording equipment and the subsequent analysis pipeline (partly manual, and partly implemented in *R*).

All the data presented in this work (apart from subsection 2.1.1) was obtained in a single test deployment of the recording equipment. This deployment was executed on a beach, where careful positioning of both the arrays and the calibration speaker was possible, and where we knew several toadfish had established nests. Two types of data were recorded. First, playbacks were produced by a *calibration speaker* and recorded by the system to correct for errors in array position and to quantify the uncertainty associated to such errors (*playback data*). Then the soundscape was recorded, which included several vocalisations by nearby toadfish males (*biological data*). The analysis pipeline outputs 2-dimensional *regions with a lower bound for credibility*, which are regions in which the true position of the fish lays with probability equal to or higher than an arbitrary value.

Chapter 2 (Materials and Methods) starts with a description of the equipment and procedure used to obtain the raw data (section 2.1), followed by a description of the pre-processing done to the recordings (section 2.2), the procedure used to measure the *TDOAs* (section 2.3) and the definition of the reference systems (section 2.4). Section 2.5 describes the model for the *TDOA* observations and how the corresponding parameters were estimated, section 2.6 explains how credible intervals for the azimuth of animals, of single vocalisations, and of the calibration speaker were obtained from those parameters and section 2.7 describes how those intervals were used to build *lower-bounded credible regions* for the positions of the fish.

Chapter 3 (Results) presents the observed distributions of *TDOA*s of stationary individuals and compares them to the estimated theoretical distributions (subsection 3.1.1), it then illustrates how the attribution of vocalizations to individuals was concordant when made from the data gathered from each of the arrays (subsection 3.1.2). Subsection 3.1.3 shows the estimated lower-bounded credible regions for the positions of the stationary individuals. Section 3.2 shows some examples of positions estimated for the sources of single sounds.

Finally, in Chapter 4 we discuss how the coherence in the results presented previously validate the system, some aspects which make this work different from previously published aporaches to fish localisation using passive acoustics, and some issues that remained unaddressed at the time of writing.

2. Materials and Methods

The recording equipment and procedure is described in section 2.1. In section 2.2 we briefly describe the pre-processing done to the recordings, and section 2.3 explains how the *TDOA*s were obtained from the pre-processed files. Section 2.4 defines the spatial reference systems used.

Section 2.5 describes how the uncertainty in the *observed TDOA*s was modelled. Contrary to similar work by others, described in the previous sections, we did not assume the *observed TDOA*s of each single vocalisation to be independent of each other, but derived a dependency structure based on the assumption of independence for *observed TOAs* (2.5.2). We also did not assume the standard deviation of the observed *TDOA* to be constant across all the experiment. Instead, we assumed it was constant for each individual in a specific location. We used the E-M algorithm to estimate the *TOA variance* for each individual male toadfish that produced vocalisations in the vicinity of the array, and then took the largest of these as a *worst-case* guess for any individual sound (2.5.3).

Section 2.6 explains how credible regions were obtained for the bearing (a) of each male identified by the E-M algorithm (2.6.1), (b) of each single vocalisation recorded (2.6.2), and (c) of the calibration speaker (2.6.3). None of these is the main output of the system. As explained in section 2.7, (a) and (b) are used to obtain *credible regions* for the position of the fish, and (c) is used to correct for error in array positioning and to include the respective uncertainty in the output, yielding *lower-bounded credible regions*. These *lower-bounded credible regions,* particularly those of *single vocalisations,* constitute the main output of the system. Such regions convey both the location of a fish in the moment it produces a vocalisation and the estimation uncertainty, without assuming the fish is stationary. The whole system is summarised in Figure 2.1, with references to the sections and subsections of this chapter as well as the *Results*.

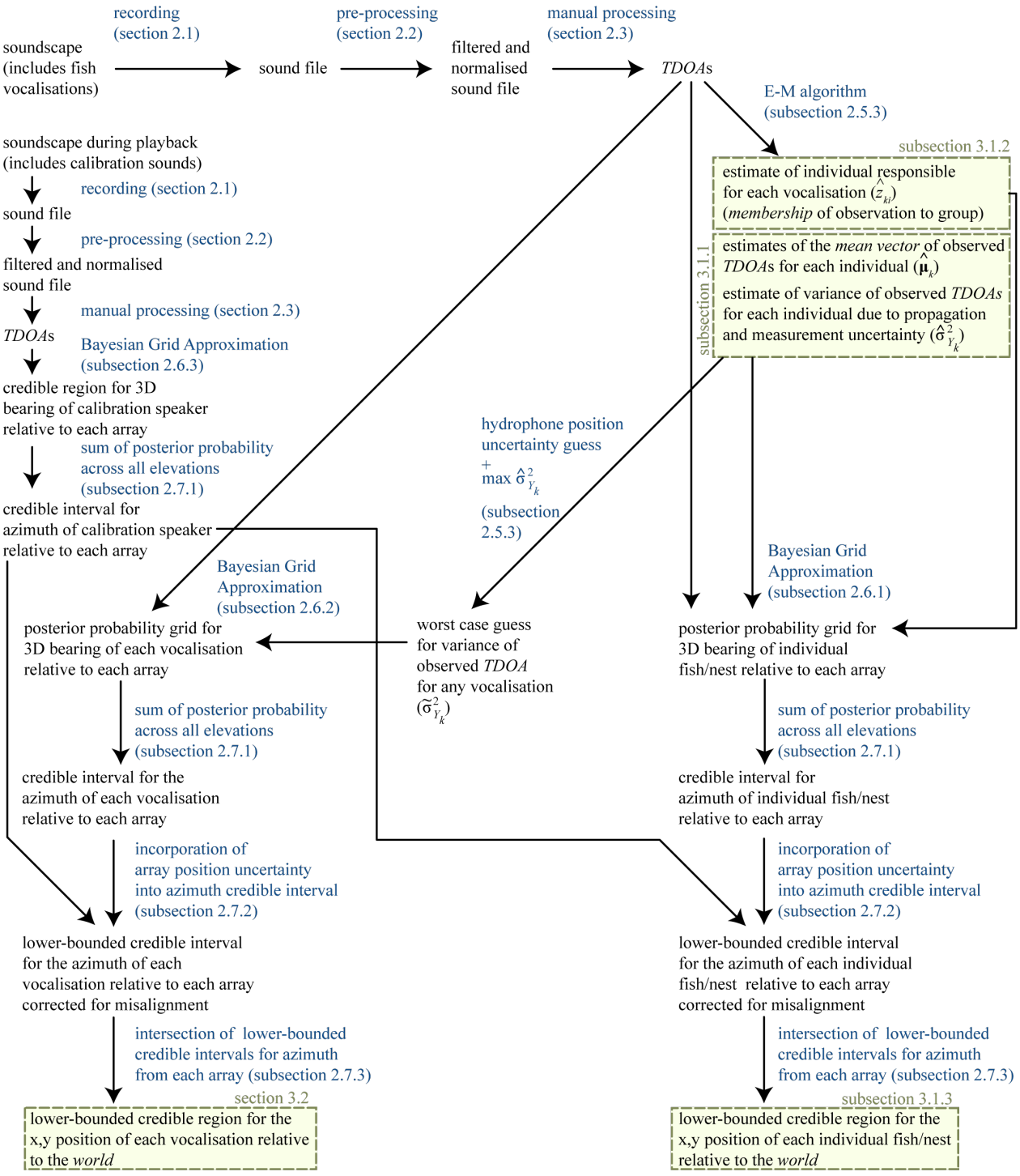

Figure 2.1 **Diagram of the system described in the present work**. References to the sections and subsections of the *Materials and Methods* where each component is detailed are included in blue text, and to those of the *Results* where each output ae marked by green boxes.

2.1.Recording equipment and deployment conditions

Two cubic steel frames (Figure 2.2) with one-meter-long edges were built of 8 mm diameter ribbed steel bars. Two diagonally opposed vertical edges extended 20 cm above the upper face of the cube. To increase

stability one bar was positioned at one diagonal of each face, three smaller bars were placed obliquely at each corner, and two were placed lateral to each of the two vertical edge extensions.

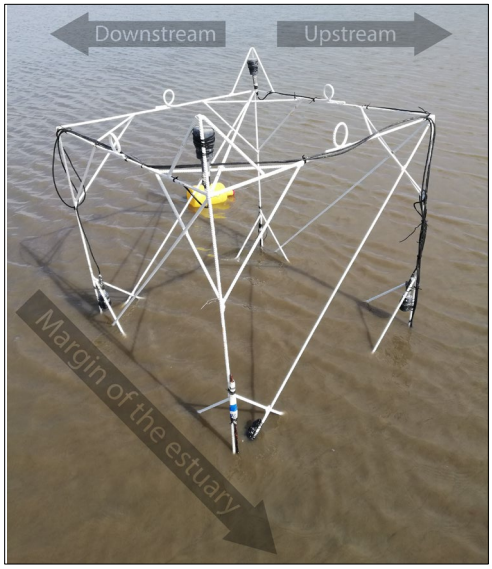

Figure 2.2 **One of the two arrays deployed.** Two of the four hydrophones can be seen just above the furthest and nearest corners of the upper face of the cube, while the other two are 15 cm above the leftmost and rightmost corners of its lower face. The pressure datalogger is partially submerged above the nearest corner of the lower face, and the upper ends of the two bars that fix the array to the bottom are visible at the nearest and furthest corners of that face. At the time the picture was taken, approximately 10 cm of the array were already submerged.

One custom made hydrophone with a piezo-electric sensor was attached to each vertical bar of the frame, 115 cm away from the bottom on the tall (120 cm long) bars, and 15 cm away from the bottom on the shorter (100 cm long) bars.

The two frames together with their respective hydrophones, will be referred to as *array 1* and *array 2*. The test deployment was done at the Portuguese Air Force Base number 6 beach in Montijo (at the south margin of the Tagus estuary – see Figure 2.6 b). Each array was positioned during low tide and kept in place by attaching its two tall vertical bars to two 80 cm long steel bars inserted about 50 cm deep into the sandy substrate. A temperature and pressure data logger (HOBO-U20-001-01, Onset Computer Corp., MA, USA) was attached to one of the diagonal bars of *array 2*, 10 cm above the bottom. The logger recorded temperature and pressure (water column height) every 5 minutes. Each array was connected to a four channel Edirol R4 digital recorder (Roland; 16 bit resolution, 96 kHz acquisition rate per channel).

The calibration sound used to correct the position of the arrays was a 20 seconds long loop of a previously recorded toadfish *boatwhistle*. It was played from a laptop running Adobe Audition 13.0, through a Technics SU-V500 Mark II amplifier to a UW30 Underwater Loudspeaker (ElectroVoice) henceforth referred to as *calibration speaker*. The omnidirectional calibration speaker was positioned about 7.5 m from each array, 10 cm from the substrate and facing up (see Figure 2.6c for the positions of the arrays and calibration speaker). The playbacks for the position calibration were made at 15h45 of July 21st 2020. At this time, the data logger registered a temperature of 23.9 $^{\circ}$ C and depth of 2.99 m. *TDOAs* of toadfish *boatwhistles* were extracted from 5 $minutes$, 40 $seconds$ of recording, starting at 15h54 of July 21st, 2020. At 15h55, the datalogger registered 23.7 $^{\circ}$ C and 3.06 m depth.

Since it was not possible to measure the salinity at the place of the deployment, the average of the values logged by the two nearest buoys from the CoastNet infrastructure (30 psu) was used. Assuming a salinity of 30 psu, the speed of sound was obtained with the function `sw_svel` from the R package `marelac` (Soetaert and Petzoldt, 2020). This function implements the *Gibbs Function*-based method described in (Feistel, 2008). The calculated speed of sound was $1526.8\ m/s$ at 15h45 and $1527.3\ m/s$ at 15h55.

2.1.1. Hydrophone selection and characterisation

For the reasons mentioned in 1.3, the custom-built piezoelectric hydrophones used in each frame were selected to have the closest frequency response possible, i.e. two groups of four matched hydrophones were selected from a total of twelve hydrophones tested (the hydrophone selection procedure was not done on site, but previously, in a pool). In order to characterize the frequency response of the hydrophones, an FFT-based transfer function was computed as the ratio of the cross spectrum between the input, an electrical signal generated by a reference hydrophone (Bruel and Kjaer 8104, frequency response 0.1 Hz–160 kHz, sensitivity −196 dB re. 1 V/µPa, conditioned by a sound level meter Bruel and Kjaer 2238 Mediator) and the output, the electrical signal of the custom-made hydrophone, to the power spectrum of the input (reference hydrophone). Both hydrophones were simultaneously activated by a sine sweep produced by a UW30 Underwater Loudspeaker, similar to the one used for calibration in the field. The loudspeaker was kept at 50 cm away from the side-by-side hydrophones and was fed by an amplifier (Technics SU-V500 Mark II). The signal to the amplifier was a sine sweep (0–20000 Hz, 41 ms) allowing the evaluation of the frequency response within this range. Moreover, the coherence function was simultaneously computed. Both functions were subjected to 25 stimuli presentations averaging. All stimuli were delivered and recorded simultaneously with a throughput rate of 50 kHz (National Instruments NI USB-6251 Multifunction I/O board). The recording (output) was synchronized (precision of ±0.01 ms at 50 kHz) relative to the stimulus (input) preventing phase change artifacts on the measurement. All functions were computed using a program made in LabVIEW (Paulo Fonseca). Figure 2.2 shows the coherence function of the eight hydrophones selected to build the arrays (panel a); the difference between the *phase frequency response* of seven of those and that of the eighth , which was arbitrarily chosen as a reference (panel b); and the *amplitude frequency response* (panel c). The hydrophones used to build the arrays were selected for being those with the most similar phase response at the frequencies that are well represented in the sounds of interest (under 500Hz) and for which a good coherence was obtained. The low coherence around 50Hz could be due to electrical noise in the recording system, or to the low gain of the hydrophones at these frequencies.

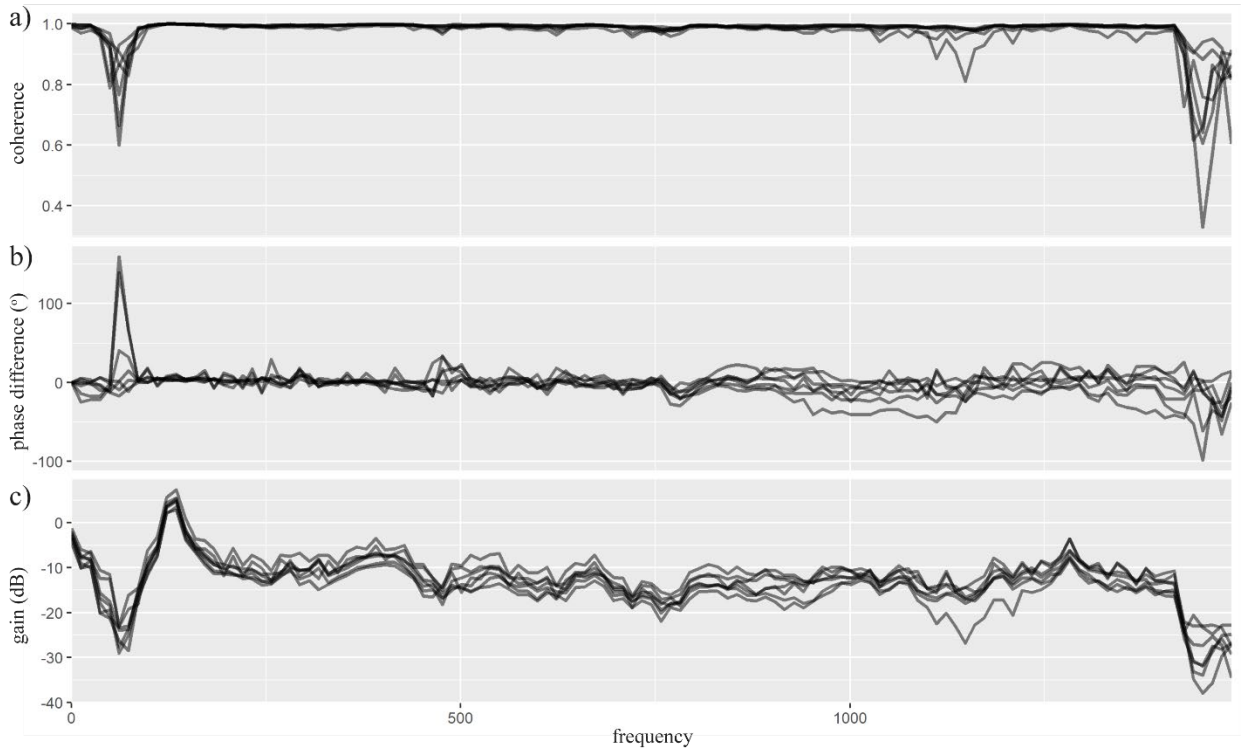

Figure 2.3 **Frequency response of the hydrophones used in the arrays.** a) Coherence of the frequency response of each the 8 hydrophones selected. b) Difference of phase frequency response between seven of the eight hydrophones selected and the eighth (arbitrarily chosen as reference), c) Amplitude frequency response of the eight hydrophones used calculated relative to a reference Bruel & Kjaer 8104 hydrophone , expressed as *gain* in dB.

2.2. Sound processing

All sound files were amplitude normalised such that the maximum recorded amplitude was stored as 90% of the maximum range, using *Adobe Audition*. They were then low pass filtered using *Raven Pro* (V1.6) batch band filter tool, excluding frequencies above 1500 Hz. *Raven Pro* uses the *Window Method* for FIR filter design (Oppenheim et al., 1999). Specifically, a Kaiser window was used with a transition bandwidth of 1.92kHz (0.02 times the Nyquist frequency) and a stop-band attenuation of 100dB.

2.3. *TOA* and *TDOA* measurements

TDOA measurements were done manually using *Raven Pro.* First, all *boatwhistles* were annotated on the sound file obtained from *array 1* (light blue areas in Figure 2.4). The resulting annotation table was used as a reference for annotation of peaks for both *array 1* and *array 2.* Then, for each sound, five 1.2 ms selections (red areas in Figure 2.4) were made such that on each selection, the same peak (positive or negative) was included in all channels. More than 5 peaks could have been used per sound. This number was chosen as a compromise between the desirable variance reduction resulting from increasing the number of data points and the consequent increased manual processing time. Raven was set up to note the moment in which the recorded pressure absolute value was maximum in each selection (*Peak Time* – red vertical lines on Figure 2.4), and the resulting annotation table was used to compute *TDOA*s. This procedure was done independently for each array.

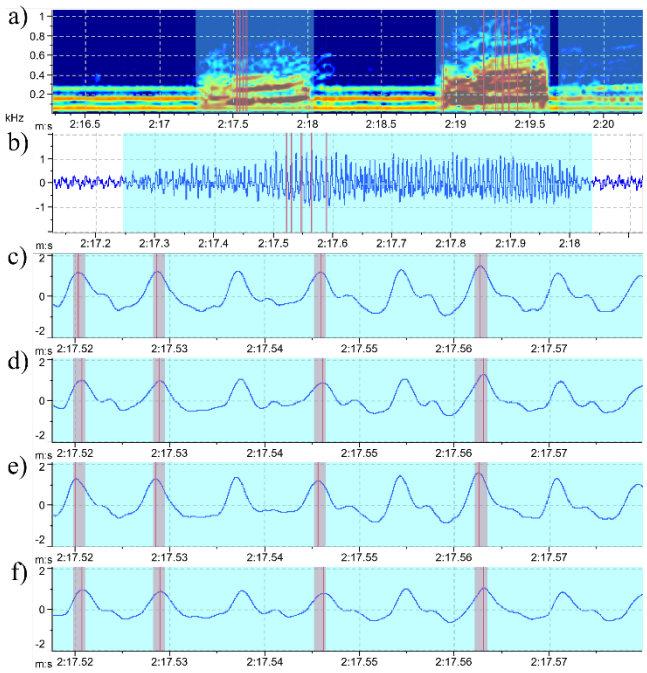

Figure 2.4 ***Raven Pro*** **spectrogram and waveform views of an example** ***boatwhistle*****.** The time scale is widest for panel a), which shows two boatwhistles with a high SNR and the beginning of a third weaker one. Panel b) spans only the duration of the first sound from panel a), and panels (c-f) show the detail of four of the five peaks whose time was measured. Panels (c-f) correspond to channels 0-3 of the sound file. Each *boatwhistle* selection is shaded in light blue, and each *peak* selection is shaded in red. The *peak time* automatically measured by the software for each *peak* selection is marked by a red vertical line.

For each of the 5 *peak selections* manually done on each sound, let the *Peak Time* on each channel be denoted t_c with $c = (0,1,2,3)$. Then one estimate for the *TDOA* between each channel 1, 2 and 3 and the reference channel (0) is

$$x_c' = t_c - t_0\,, \qquad c = (1,2,3)\,. \tag{2.1}$$

As noted previously, 5 selections were made per sound, and therefore 5 such estimates could be obtained from each vocalisation. Yet, since the observations of peak times from the same vocalisation are unlikely to be independent, and since it is not easy to estimate their dependency structure, they were aggregated as a median, and the latter was considered the elementary observation of the experiment. That is to say, each median of the five $\boldsymbol{x'} = (x'_1, x'_2, x'_3)$ obtained for one vocalisation was considered a realization of the random vector $\boldsymbol{X}$ presented in section 2.5.

Of a total 264 toadfish vocalisations manually detected in the recording, this procedure was done (and observed *TDOAs* obtained) for a total 185 sounds. Of these, 66 had their *TDOA*s obtained relative to the hydrophones of both arrays, 42 only relative to those of array 1, and 77 only relative to those of array 2. Whenever it was not possible to obtain the *TDOA*s of a sound relative to one or both arrays, this was due to distortions in the signal due low signal-to-noise ratio or to interference caused by other sounds in the environment.

2.4. Reference systems

2.4.1. Array reference system

The bearing of a sound is defined in this work as the pair of angles (α , θ) that defines the direction of a sound arriving at an array, such that the axes are parallel to the edges of the cubic frame supporting the hydrophones; α (the azimuth) is measured counter-clockwise (as seen from above) over the $x0y$ plane from the x axis; and θ is measured from $x0y$ plane over the vertical plane that includes the source of the sound and the origin of the reference system (see Figure 2.5). The origin of the reference system is at the centre of the array.

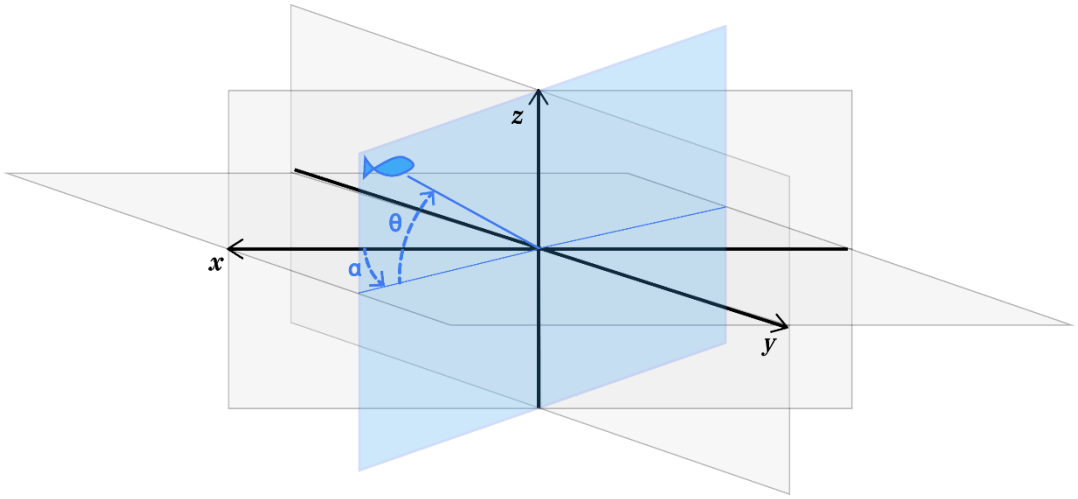

Figure 2.5 **Array reference system (for bearing estimation).** The origin of the reference system corresponds to the position of the array.

2.4.2. World reference system

Since there is an *array reference system* relative to each array, when crossing azimuths to obtain positions in the horizontal plane, we must define a reference system that is independent of the array positions. The coordinates of that system will be denoted x_w, y_w, z_w and it is defined as follows (see Figure 2.6):

- The origin is midway between the 2 array centres, at approximately 38°41'32"N, 9° 2'59"W;
- x_w is directed horizontally, away from the left margin of the estuary;
- y_w is directed horizontally, downstream; and
- z_w is directed upwards.

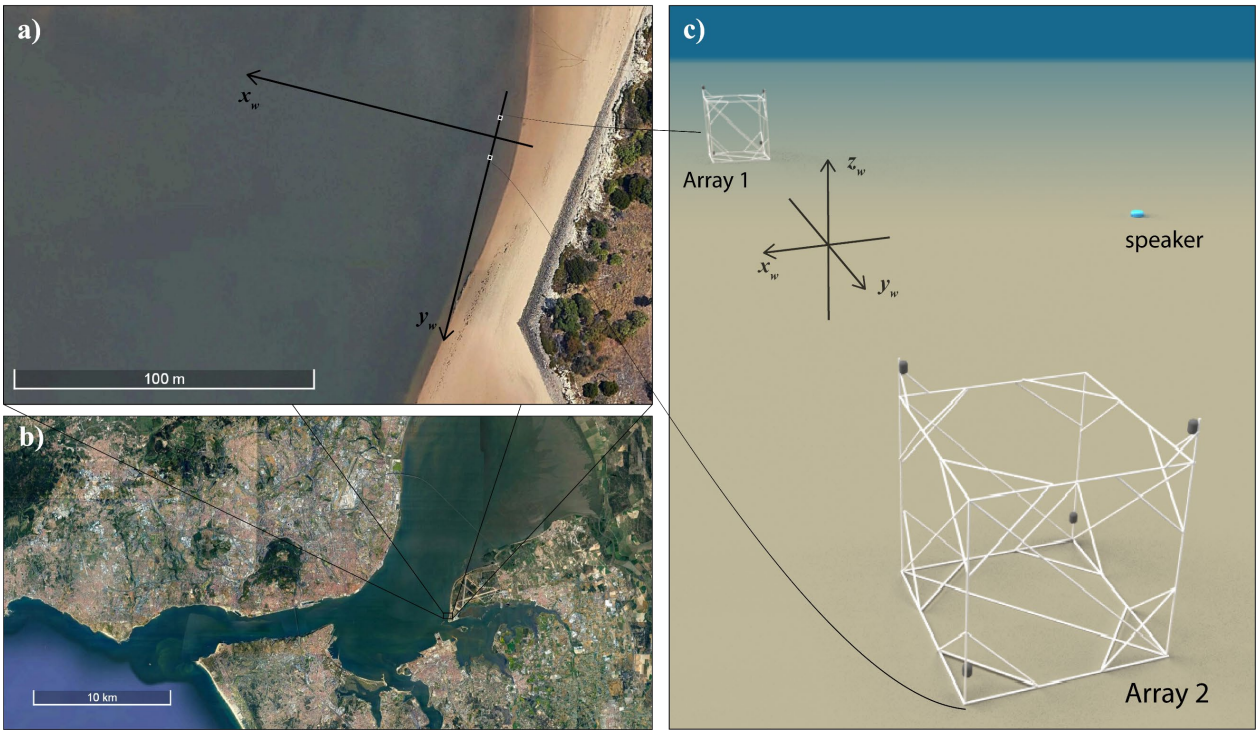

Figure 2.6 **Position of the arrays on the beach at BA6.** a) World reference system (for position estimation) – the z_w axis is not shown as it is orthogonal to the plane of the picture; b) overview of the estuary; c) position of the arrays relative to the calibration speaker. The arrays are on the y axis at positions -6m (array 1) and +6m (array 2).

The arrays were placed parallel to the beach (i.e. over the y_w axis), and the origin of the (x_w, y_w, z_w) was defined so that the positions of the arrays *1* and *2* were $(0, -6, 0)$ and $(0, 6, 0)$, respectively, thereby establishing the correspondence between the origins of each *array reference system* and the *world reference system*. The bottom of the estuary in the place of deployment was considered approximately horizontal and since the frame was placed on the bottom, $x0y$ coincides with $x_w 0 y_w$, and the direction of z coincides with the direction of z_w.

An attempt was made during deployment to have directions of x and y correspond to those of x_w and y_w, respectively. If it had been possible to do so, the positions of the sensors relative to the *World reference system* would be as in Table 1. Yet, it was not possible to perfectly align the arrays in place, and the difference (rotation around the z axis) between each *array reference system* and the *world reference system* was estimated using a *calibration speaker* positioned at coordinates $(-4.5, 0, -0.55)m$ relative to the world reference system, as described in 2.7.2.

Table 1 - planned positions of the hydrophones relative to the World reference system.

Array	Hydrophone	x_w	y_w	z_w
1	0	-0.5m	-5.5m	0.5m
	1	0.5m	-6.5m	0.5m
	2	-0.5m	-6.5m	-0.5m
	3	0.5m	-5.5m	-0.5m
2	0	-0.5m	6.5m	0.5m
	1	0.5m	5.5m	0.5m
	2	-0.5m	5.5m	-0.5m
	3	0.5m	6.5m	-0.5m

2.5. Modelling TDOAs

The measured *TDOAs* of a sound to each of the three non-reference hydrophones relative to an arbitrary reference hydrophone (X_1, X_2, X_3) are modelled as a trivariate Normal distribution

$$\boldsymbol{X} \frown Normal_3(\boldsymbol{\mu}, \boldsymbol{\Sigma}) , \tag{2.2}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are as described in the following sections.

2.5.1. Mean vector

Each element of the *mean* vector of this distribution is a function of the bearing of the source via a geometric and physical model

$$\mu_i(\boldsymbol{b}) = \frac{d_i}{C} = \frac{1}{C}(\boldsymbol{r}_i \cdot \boldsymbol{u}) = \frac{1}{C}(\boldsymbol{r}_i \cdot (-\boldsymbol{b})), \qquad i = (1,2,3) \tag{2.3}$$

where d_i is the distance travelled by the wave front from the moment it crosses a *non-reference hydrophone* (HP_i) to the moment it crosses the *reference hydrophone* (HP_0); $\boldsymbol{u}$ is a unitary vector representing the direction of sound propagation; $\boldsymbol{b}$ is also a unitary vector, representing the bearing of the source relative to the array; $\boldsymbol{r}_i$ represents the position of HP_i relative to HP_0; and C is the speed of sound in the water.

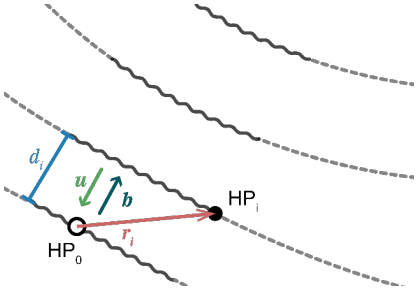

Figure 2.7 **Schematic representation of the geometry involved in the model described by equation** (1.2)**.** $\boldsymbol{d_i}$ (the distance travelled by the wave front from the moment it crosses HP_i to the moment it crosses HP_0) is directly proportional to $\boldsymbol{\mu_i}$ (the expected $TDOA_i$) through the constant C (speed of sound). For simplicity, only one pair of hydrophones is shown.

Note that $\boldsymbol{u}$ represents the direction of propagation of sound from the moment the wave front crosses one hydrophone to the moment it crosses another. Therefore, $\boldsymbol{b}$ is only the bearing of the source if the wavefront travelled in a straight line from the source to the array. If it changed direction due to reflection or refraction, then $\boldsymbol{b}$ will not in general be the true bearing of the source.

On the other hand, if the sound got reflected only off the surface and/or bottom, and these are assumed horizontal, then the azimuth component of $\boldsymbol{b}$ will correspond to the true azimuth of the source, even though the same might not be true for the elevation component (see Figure 1.2).

2.5.2. Covariance matrix

The variability of (X_1, X_2, X_3) has 3 independent sources: variation in the *TOA* due to propagation variability, uncertainty in the *TOA* measurements, and uncertainty in the hydrophones positions.

Let the *TOA* of the sound to each hydrophone HP_i be denoted Y_i, and the *TOA* of the sound to the reference hydrophone be denoted Y_0. Then,

$$X_i = Y_i - Y_0, \quad i = (1,2,3). \tag{2.4}$$

Clearly the different *TDOA* observations are not independent since they have Y_0 in common. Assuming the *observed TOAs* of a given sound (Y_0, Y_1, Y_2, Y_3) are independent and have the same variance σ_Y^2, the covariance structure of (X_1, X_2, X_3) is specified at the expense of σ_Y^2 by

$$\boldsymbol{\Sigma}(\sigma_Y^2) = \begin{bmatrix} 2\,\sigma_Y^2 & \sigma_Y^2 & \sigma_Y^2 \\ \sigma_Y^2 & 2\sigma_Y^2 & \sigma_Y^2 \\ \sigma_Y^2 & \sigma_Y^2 & \sigma_Y^2 \end{bmatrix}. \tag{2.5}$$

Furthermore, as noted previously, σ_Y^2 is due to 3 independent sources of variation, and may therefore be written

$$\sigma_Y^2 = \sigma_{prop}^2 + \sigma_{meas}^2 + \sigma_{hp}^2 = \sigma_{prop+meas}^2 + \sigma_{hp}^2 \tag{2.6}$$

where σ_{prop}^2 is the variance due to propagation uncertainty; σ_{meas}^2 is the variance due to measurement uncertainty and σ_{hp}^2 is the variance due to hydrophone position uncertainty.

$\sigma_{prop}^2 + \sigma_{meas}^2$ were combined into $\sigma_{prop+meas}^2$ because this combined value may be estimated from data as explained in 2.5.3.

2.5.3. *Observed TOA* uncertainty (σ_Y^2)

As noted in 1.5.2, to reduce the dimensionality of the estimation problem, Nosal and Frazer (2007) work with the likelihood surfaces conditional on *worst case* values for the dispersion parameters. In a similar fashion, but under the Bayesian paradigm, we maximise and integrate over the posterior probability as a function of the azimuth and elevation (the location parameters), conditional on a worst-case value for *TOA variance* (the dispersion parameter).

Accordingly, we define the worst case value for σ_Y^2 on which to condition the likelihood of (X_1, X_2, X_3) to be

$$\tilde{\sigma}_Y^2 = \tilde{\sigma}_{prop+meas}^2 + \tilde{\sigma}_{hp}^2 \,. \tag{2.7}$$

$\tilde{\sigma}_{hp}^2$ is obtained from a somewhat subjective but conservative guess of the standard deviation ($\tilde{\sigma}_{\text{pos}} = 2\ cm$) of the hydrophones position in the array:

$$\tilde{\sigma}_{hp}^2 = \left(\frac{1}{C}\,\tilde{\sigma}_{\text{pos}}\right)^2 . \tag{2.8}$$

This guess was based on the perceived accuracy of positioning of the array frame elements while the frame was built, as well as on measurements of the distances between hydrophones after they were mounted onto the frame.

Because male toadfish are known to vocalise from fixed positions to attract females, we can assume the n toadfish vocalisations recorded by an array during a short period of time come from a finite number (r) of positions in space, each corresponding to one individual male (k), and having a potentially unique bearing ($\boldsymbol{b_k}$, $\boldsymbol{k} = 1, 2, \ldots, r$) relative to the array. That being so, each observed vector of *TDOA*s ($\boldsymbol{x}_i$) obtained from the *biological data* is a realisation of a random vector taken from a mixture of r trivariate Normal distributions each with mean vector $\boldsymbol{\mu}_k$ as defined by (2.3) and covariance matrix $\boldsymbol{\Sigma}\left(\sigma_{\hat{Y}_k}^2\right)$ as defined by (2.5). Taking advantage of that knowledge to obtain a conservative guess for variance due to propagation and measurement uncertainty ($\tilde{\sigma}_{prop+meas}^2$), we estimated $\boldsymbol{\mu_k}$ and $\sigma_{\hat{Y}\,k}^2$ for each group k (possibly an individual male). For that, we used an *Expectation-Maximization (E-M) algorithm* (e.g. Friedman et al., 2001; Murteira and Antunes, 2012). Since each observation i belongs to one of r Normal distributions, the unobserved variable z_{ki} is defined as the *membership* of each observation i to the distribution (or group) k. This variable takes the value 1 if the observation i belongs to group k or 0 otherwise. Yet, if we look at z_{ki} as the realization of a random variable Z_{ki}, it's expected value may take any value from 0 to 1.

The E-M algorithm estimates (at each iteration j) the *membership* of each observation $\boldsymbol{x}_i$ to each group k as

$$z_{ki}^{(j)} = E\left(Z_{ki}|\boldsymbol{X}, \boldsymbol{p}^{(j-1)}, \boldsymbol{M}^{(j-1)}, \boldsymbol{\sigma}_{\hat{Y}}^{2\,(j-1)}\right) = \frac{f\left(\boldsymbol{x}_i; \boldsymbol{\mu_k}^{(j-1)}, \sigma_{\hat{Y}\,\boldsymbol{k}}^{2\,(j-1)}\right) \cdot p_k^{(j-1)}}{\sum_{l=1}^{r} f\left(\boldsymbol{x}_i; \boldsymbol{\mu_l}^{(j-1)}, \sigma_{\hat{Y}\,\boldsymbol{l}}^{2\,(j-1)}\right) \cdot p_l^{(j-1)}}, k = (1,2, \ldots, r) \tag{2.9}$$

with $\boldsymbol{X}$ being the data; $\boldsymbol{p}^{(j-1)}$ a vector of r elements $\boldsymbol{p_k}^{(j-1)}$; $\boldsymbol{p_k}^{(j-1)}$ the previous iteration estimates for the proportion of sounds belonging to group k; $\boldsymbol{M}^{(j-1)}$ a $(3 \times k)$ matrix, each column the previous iteration estimate for the mean vector of group k ($\boldsymbol{\mu_k}^{(j-1)}$); $\boldsymbol{\sigma_Y^{2\,(j-1)}}$ a vector, each of its r values $\sigma_{\hat{Y}\,\boldsymbol{k}}^{2\,(j-1)}$ the estimate of $\sigma_{\hat{Y}}^{2}$ for group k; $\boldsymbol{x}_i$ the ith observation of the vector (X_1, X_2, X_3); and $f\left(\boldsymbol{x}_i; \boldsymbol{\mu_k}^{(j-1)}, \sigma_{\hat{Y}\,\boldsymbol{k}}^{2\,(j-1)}\right)$ the probability density function of a trivariate Normal distribution with mean $\boldsymbol{\mu_k}^{(j-1)}$ and covariance matrix $\boldsymbol{\Sigma}(\sigma_{\hat{Y}\,\boldsymbol{k}}^{2\,(j-1)})$ (see (2.5)) evaluated at the point $\boldsymbol{x}_i$.

Each parameter estimate is then replaced by its maximum likelihood estimate conditional on $z_{ki}^{(j)}$ and the most recent estimate for all other parameters:

$$\boldsymbol{p_k^{(j)}} = \frac{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} z_{ki}^{(j)}}{\boldsymbol{n}} \tag{2.10}$$

$$\boldsymbol{\mu_k^{(j)}} = \frac{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \boldsymbol{x}_i \, z_{ki}^{(j)}}{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} z_{ki}^{(j)}} \tag{2.11}$$

$$\sigma_{Y\boldsymbol{k}}^{2\,(j)} = \frac{\mathbf{1}}{\mathbf{12}} \frac{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} z_{ki}^{(j)} \left[3 \cdot \sum_{g=1}^{3} \left(x_{ig} - \mu_{gk}^{(j)} \right)^2 - 2 \cdot \sum_{g=1}^{2} \sum_{h=g}^{3} \left(x_{ig} - \mu_{gk}^{(j)} \right) \left(x_{ih} - \mu_{hk}^{(j)} \right) \right]}{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} z_{ki}^{(j)}} \tag{2.12}$$

$$, k = (1, 2, \ldots, r) \, .$$

To choose $\boldsymbol{p^{(0)}}, \boldsymbol{M^{(0)}}, \boldsymbol{\sigma_Y^{2\,(0)}}$ we performed a *k-means* clustering of the data, with the default method implemented in *R stats* package (Hartigan and Wong, 1979). The clustering was repeated 100 times with different random initial cluster centres and the results of the run with minimum *within cluster sum of squares* was used. The *proportion of observations allocated to each cluster*, the *centres of the clusters* and the *within cluster sum of squares* were used for the initialisation values $\boldsymbol{p^{(0)}}, \boldsymbol{M^{(0)}}, \boldsymbol{\sigma_Y^{2\,(0)}}$, respectively.

The iterations were stopped whenever all parameters remained *nearly equal* from one iteration to the next. *Near equality* was evaluated with the *R* function `all.equal()` which evaluates `TRUE` when the difference between the numbers to compare is smaller than $\sqrt{\delta}$, where δ is the smallest number for which the computer can distinguish 1 and $1 + \delta$.

The estimates $\hat{\boldsymbol{\mu}}_k$, $\hat{\sigma}_{Y\,k}^2$ and $\hat{\boldsymbol{z}}_k$ (for $\boldsymbol{\mu_k}$, $\sigma_{Y\,k}^2$ and $\boldsymbol{z}_k$) are then the values of $\boldsymbol{z_k^{(j)}}$, $\boldsymbol{\mu_k^{(j)}}$ and $\sigma_{Y\boldsymbol{k}}^{2\,(j)}$ at the last iteration. Note that the elements $(\hat{z}_{ki})$ of the vector $\hat{\boldsymbol{z}}_k$ correspond to the *estimated membership* of each sound to a group of sounds, and even though the actual (unobserved) *membership* (z_{ki}) of each observation i to group k is either 0 or 1, its estimate is only constrained – by equation (2.9) – to be between 0 and 1 and to have $\sum_k z_{ki} = 1$ for each observation.

By listening to the recording and visually inspecting the spectrogram and oscillogram it is possible to recognise the vocalisations emitted by different individuals (Amorim and Vasconcelos, 2008). Yet, the number of clusters r was chosen by inspecting the histograms of the observed *TDOAs* and choosing for the highest number of modes among the 3 pairs of hydrophones of each array, so that aural and visual inspection of the sound could work as a validation of the method (see section 3.2).

After running the E-M algorithm, the histograms of X_1, X_2 and X_3 (the *TDOAs* of each pair of hydrophones) weighted by each vector $\hat{\boldsymbol{z}}_k$ were analysed (Figure 3.1 and Figure 3.2), and whenever several modes were apparent on any of the $3r$ histograms the algorithm was repeated with r incremented by 1.

With r selected as described above, the E-M algorithm always converged in less than 50 iterations yielding one group with few observations and higher dispersion, and several groups with observed *TDOA*s, apparently consistent with being realizations of multivariate normal distributions (see 3.1.1). We assumed the group with fewer observations was simply aggregating sounds that came from (possibly multiple) individuals which produced few vocalisations during the data collection period.

When estimating bearings or positions for single vocalisations without assuming they belonged to a group emitted from the same position (2.6.2), the highest estimate of $\sigma^2_{Y_k}$ obtained with the E-M algorithm ($\widehat{\sigma}^2_{Y_k}$) of all groups excluding that with the least observations was used as a conservative guess for variance due to propagation and measurement uncertainty

$$\tilde{\sigma}^2_{prop+meas} \equiv \max_k\left(\widehat{\sigma}^2_{Y_k}\right), \qquad \left\{k:\ k \in \{1,2,\dots,r\} \wedge \sum_{i=1}^{n} \hat{z}_{\boldsymbol{ki}} > min\left(\sum_{i=1}^{n} \hat{z}_{\boldsymbol{ki}}\right)\right\} . \tag{2.13}$$

On the other hand, when estimating bearings or positions of individuals assumed to have produced multiple vocalisations from a single location (2.6.1), the corresponding *TOA* variance estimate ($\widehat{\sigma}^2_{Y_k}$) was used.

2.6. Estimating Bearings

This section describes how the *observed TDOA*s were used to obtain credible regions (a) for the bearings of the individual stationary males that produce multiple vocalisations (2.6.1), (b) for the bearings of each vocalising fish without assuming it produces multiple vocalisations from a single position (2.6.2), and (c) for the bearing of the calibration speaker relative to each array.

The reduced dimensionality of the problem that resulted from conditioning the likelihood on the worst case guesses, allows the use of a simple yet computationally intensive numerical approximation of the posterior probability, namely the *Bayesian grid approximation* (McElreath, 2020).

With this approach, the posterior probability of (α, θ) being in the cell of an $r \times s$ grid centred at (α_i, θ_j) is approximated by

$$\mathrm{P}(\alpha_i, \theta_j|\boldsymbol{x}) = \frac{L(\boldsymbol{X}; \alpha_i, \theta_j) \cdot \mathrm{P}(\alpha_i, \theta_j)}{\sum_{k=1}^{r}\sum_{l=1}^{s} L(\boldsymbol{X}; \alpha_k, \theta_l) \cdot \mathrm{P}(\alpha_k, \theta_l)} \tag{2.14}$$

with $i = (1,2,\dots,r);\ j = (1,2,\dots,s)$;$L(\boldsymbol{X}; \alpha_i, \theta_j)$ the likelihood of the observed values $\boldsymbol{X}$ if $\alpha = \alpha_i$ and $\theta = \theta_j$**;** and $P(\alpha_i, \theta_j)$ the *a priori* probability of (α, θ) being in a cell centred at (α_i, θ_j).

This numerical method requires few assumptions and is mostly limited by the fact that it scales poorly with the dimensionality of the parameters. Indeed, the only underlying assumption is that the grid is fine enough for the *posterior density* of (α_i, θ_j) to be representative of the posterior density of all the points in the cell centred thereat.

Since for any $\boldsymbol{b_{ij}} = (\alpha_i, \theta_j)$ there is a $\boldsymbol{\mu}(\boldsymbol{b_{ij}}) = (\mu_1, \mu_2, \mu_3)$ defined by (2.3),we can rewrite (2.14) as

$$\mathrm{P}(\alpha_i, \theta_j|\boldsymbol{X}) = \frac{L\left(\boldsymbol{X}; \boldsymbol{\mu}(\alpha_i, \theta_j)\right) \cdot \mathrm{P}(\alpha_i, \theta_j)}{\sum_{k=1}^{r}\sum_{l=1}^{s} L\left(\boldsymbol{X}; \boldsymbol{\mu}(\alpha_k, \theta_l)\right) \cdot \mathrm{P}(\alpha_k, \theta_l)} . \tag{2.15}$$

The *highest posterior density* (HPD) credible region with probability p is then approximated by ordering the pairs (α_i, θ_j) by decreasing $\mathrm{P}(\alpha_i, \theta_j|\boldsymbol{X})$ and selecting those whose cumulative sum is less than or equal to p.

$L(\boldsymbol{X};\ \boldsymbol{\mu}(\alpha_i, \theta_j))$ will have different expressions depending on the problem. Specifically, in the next sections we defined it as being the likelihood of several observations from a mixture of trivariate Normal

distributions (2.6.1), the likelihood of a single observation of a trivariate Normal distribution (2.6.2), and the likelihood of several observations from a single trivariate Normal distribution (2.6.3).

2.6.1. Bearings of individuals/nests

To estimate credible regions for the bearings of each individual vocalising from a nest (corresponding to each group k except that with the least number of observations), we replaced $L\left(\boldsymbol{X};\boldsymbol{\mu}(\alpha_i,\theta_j)\right)$ in (2.14) with the likelihood of the bearing of each cell of the grid (α_i,θ_j) $for\ the\ n$ observed triplets of *TDOA*s, conditional on the clustering done by the E-M algorithm ($\hat{\boldsymbol{z}}_k$), on $\sigma^2_{prop+meas}=\hat{\sigma}^2_{Y_k}$ and on $\sigma^2_{hp}=\tilde{\sigma}^2_{hp}$:

$$L_k\left(\boldsymbol{X};\boldsymbol{\mu}(\alpha_i,\theta_j)|\hat{\boldsymbol{z}}_k;\ \sigma^2_{prop+meas}=\hat{\sigma}^2_{Y_k};\ \sigma^2_{hp}=\tilde{\sigma}^2_{hp}\right)=\prod_{l=1}^{n} f\left(\boldsymbol{x}_l;\boldsymbol{\mu}(\alpha_i,\theta_j),\hat{\sigma}^2_{Y_k}+\tilde{\sigma}^2_{hp}\right)^{\hat{z}_{kl}}\ , \qquad (2.16)$$

where $f\left(\boldsymbol{x}_l;\boldsymbol{\mu}(\alpha_i,\theta_j),\hat{\sigma}^2_{Y_k}+\tilde{\sigma}^2_{hp}\right)$ is the probability density function of a trivariate Normal distribution with mean $\boldsymbol{\mu}(\alpha_i,\theta_j)$ and covariance matrix $\boldsymbol{\Sigma}(\hat{\sigma}^2_{Y_k}+\tilde{\sigma}^2_{hp})$ (see (2.5)) evaluated at the point $\boldsymbol{x}_l$; $\hat{\sigma}^2_{Y_k}$ is the E-M algorithm estimate of $\sigma^2_{Y_k}$, and $\tilde{\sigma}^2_{hp}$ is as defined in (2.4).

2.6.2. Bearings of single observations

The approach presented in 2.6.1 leaves out the data from fish that failed to vocalise enough during data collection to produce a cluster that the E-M algorithm could resolve. It is also limited to animals that remain in the same location during that period.

In order to obtain a credible region for each sound detected that did not depend on the clustering done by the E-M algorithm, we used the likelihood of the bearing of each cell (α_i,θ_j) of the grid given the observed vector ($\boldsymbol{x}$) of *TDOAs*, conditional on $\tilde{\sigma}^2_Y$ as defined in (2.7):

$$L\left(\boldsymbol{x};\boldsymbol{\mu}(\alpha_i,\theta_j)|\tilde{\sigma}^2_Y\right)=f\left(\boldsymbol{x};\boldsymbol{\mu}(\alpha_i,\theta_j),\tilde{\sigma}^2_Y\right)\ , \qquad (2.17)$$

where $f\left(\boldsymbol{x};\boldsymbol{\mu}(\alpha_i,\theta_j),\tilde{\sigma}^2_Y\right)$ is the probability density function of a trivariate Normal distribution with mean $\boldsymbol{\mu}(\alpha_i,\theta_j)$ and covariance matrix $\boldsymbol{\Sigma}(\tilde{\sigma}^2_Y)$ (see (2.5)) evaluated at the point $\boldsymbol{x}$.

2.6.3. Bearing of calibration speaker

The main outputs of the system are 2-dimensional credible regions for the position of vocalising fish in the horizontal plane. These were obtained, as exposed in section 2.7, by crossing credible intervals for the azimuth of a sound or group of sounds relative to 2 arrays. To do so, both the position on the horizontal plane and the relative orientation of the arrays had to be known. Ideally these would be known with a negligible error by careful placement during deployment. Yet, we found that using a single calibrating speaker to produce sounds that were then used to calibrate the orientation of the arrays prevented a significant amount of bias due to imperfect placement of the devices and allowed to quantify the uncertainty associated to array orientation (see 2.7.2 for details).

To perform this calibration, the credible region for the bearing of the speaker relative to each array was obtained with $L\left(\boldsymbol{X};\boldsymbol{\mu}(\alpha_i,\theta_j)\right)$ in (2.15) being replaced by the likelihood of the bearing of each cell of the grid (α_i,θ_j) for the n observed triplets of *TDOA*s, conditional on $\sigma^2_{prop+meas}=\check{\sigma}^2_{Y_k}$ and on $\sigma^2_{hp}=\tilde{\sigma}^2_{hp}$:

$$L\left(\boldsymbol{X};\boldsymbol{\mu}\left(\alpha_i,\theta_j\right)|\sigma^2_{prop+meas}=\breve{\sigma}^2_{\dot{Y}k};\ \sigma^2_{hp}=\tilde{\sigma}^2_{hp}\right)=\prod_{l=1}^{n} f\left(\boldsymbol{x}_l;\boldsymbol{\mu}\left(\alpha_i,\theta_j\right),\breve{\sigma}^2_{\dot{Y}k}+\tilde{\sigma}^2_{hp}\right)\ \ , \qquad (2.18)$$

where

$$\breve{\sigma}^2_{\dot{Y}k}=\frac{\sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}}\left[3\cdot\sum_{g=1}^{3}\left(x_{ig}-\mu_g\left(\alpha_i,\theta_j\right)\right)^2\ -\ 2\ \cdot\sum_{g=1}^{2}\sum_{h=g}^{3}\left(x_{ig}\ -\ \mu_g\left(\alpha_i,\theta_j\right)\right)\left(x_{ih}\ -\ \mu_h\left(\alpha_i,\theta_j\right)\right)\right]}{12\ n} \qquad (2.19)$$

is the maximum likelihood estimator of $\sigma^2_{\dot{Y}k}$ from the speaker data (which is known to be in a fixed position).

2.6.4. The prior

In the context of the Bayesian grid approximation, the prior is the *a priori* probability mass of each grid cell centred at a bearing (α_i,θ_j).

If that grid is regular (having constant $\delta_\alpha=\alpha_{i+1}-\alpha_i$ and $\delta_\theta=\theta_{i+1}-\theta_i$), and assuming the sound may arrive from any direction with equal probability, then the probability mass of each grid cell will decrease as the elevation departs from 0° (both towards 90° and -90°).

To visualise this, one can picture the sound traveling as a ray towards the hydrophone array at the centre of a sphere (Figure 2.8). Saying the sound may arrive from any direction with equal probability is the same as saying it may cross the spheric surface at any point with equal probability. Now visualise the bearing grid cells as the rectangular pyramids with apex at the centre of the sphere (also the position of the array), height equal to the sphere radius, and lateral face orientations defined by the limits of the grid cells $(\alpha_i\pm\delta_\alpha/2\ ,\theta_j\pm\delta_\theta/2)$. Their base approximates the area of spheric surface that corresponds to the cell. Since all points of the spheric surface have equal (prior) probability, each cell has probability approximated by the relative area of the base of the corresponding pyramid

$$\mathrm{P}\left(\alpha_i,\theta_j\right)=\frac{2\ r\ tg(\delta_\theta/2)\cdot 2\ r\ cos\left(\theta_j\right)\ tg(\delta_\alpha/2)}{\sum_{k=1}^{r}\sum_{l=1}^{s}2\ r\ tg(\delta_\theta/2)\cdot 2\ r\ cos(\theta_l)\ tg(\delta_\alpha/2)}=\frac{cos\left(\theta_j\right)}{r\ \sum_{l=1}^{s}cos(\theta_l)}\ . \qquad (2.20)$$

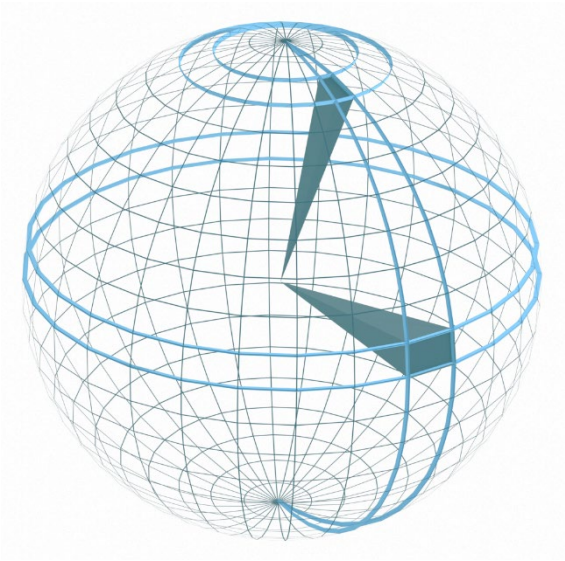

Figure 2.8 **Prior grid cells visualized as pyramid bases.** Two pyramids with shared apex, equal height and lateral faces defined by the margins of two different cells of the bearings grid are highlighted. It is apparent that to a higher elevation corresponds a lower pyramid base area. This is despite the angles (δ_α and δ_θ) between opposite lateral faces being the same in both cases.

2.7. Positions from crossed azimuths

In this section, we expose the process of obtaining 2-dimensional *lower-bounded credible regions* for the *location* of the fish from the credible regions for the *bearing* of the fish relative to each array. This involves reducing the dimensionality of the bearing credible regions by computing the HPD credible interval of the *azimuth* relative to each array (2.7.1), estimating the uncertainty regarding the array's position (2.7.2), incorporating that uncertainty into the credible region for the azimuth (2.7.3), and intersecting the resulting *lower-bounded credible regions* (2.7.4).

2.7.1. Azimuth Credible Interval from Bearing Credible Region

Once (2.15) has been evaluated for all cells in the 2-dimensional (α_i, θ_j) grid, it is easy to obtain a grid approximation of the marginal posterior probability of the azimuth

$$\mathrm{P}(\alpha_i|\boldsymbol{x}) = \sum_{j=1}^{s} \mathrm{P}(\alpha_i, \theta_j|\boldsymbol{x}) \quad . \tag{2.21}$$

Obtaining an HPD credible interval of probability p is then a matter of ordering by decreasing $\mathrm{P}(\alpha_i|\boldsymbol{x})$ and selecting those α_i with cumulative sum ($s_c(\alpha_i)$) lower than p. As long as (2.21) has a single peak, that interval may be specified by its upper (α_{up}) and lower (α_{low}) limits, which are computed as

$$\alpha_{low} = \begin{cases} \min(\alpha_i) - \delta_\alpha/2 : s_c(\alpha_i) < p & if\ [(\max(\alpha_i) - \min(\alpha_i)) : s_c(\alpha_i) < p] < 180^\circ \\ \max(\alpha_i) - \delta_\alpha/2 : s_c(\alpha_i) < p & if\ [(\max(\alpha_i) - \min(\alpha_i)) : s_c(\alpha_i) < p] > 180^\circ \end{cases} \quad (2.22)$$

and

$$\alpha_{up} = \begin{cases} max(\alpha_i) + \delta_\alpha/2 : s_c(\alpha_i) < p & if\ [(\max(\alpha_i) - \min(\alpha_i)) : s_c(\alpha_i) < p] < 180^\circ \\ min(\alpha_i) + \delta_\alpha/2 : s_c(\alpha_i) < p & if\ [(\max(\alpha_i) - \min(\alpha_i)) : s_c(\alpha_i) < p] > 180^\circ \end{cases} . \quad (2.23)$$

This definition of lower and upper limits assumes all intervals are less than 180° wide, and it means the angle rotates counter-clockwise from the lower to the upper limit regardless of which limit is numerically larger. That is, if $\alpha = 0$ is in the interval, the lower limit is between 180° and 360° and the upper limit is between 0° and 180°.

2.7.2. Uncertainty in array orientation

As alluded to in 2.4.2, the orientation of the arrays relative to the *world reference system* is not known but estimated.

Knowing that $x0y$ coincides with x_w0y_w, this orientation may be fully specified by the angle between the directions of the axes x and x_w, or equivalently those of the axes y and y_w.

Let α'_{sp} be the angle measured over the plane x_w0y_w from the line parallel to the x_w axis that crosses the centre of the array to the line that connects the centre of the array to the position of the calibration speaker. Also let α_{sp} be the azimuth of the speaker relative to the *array reference system* as defined in 2.4. Note that if the array and world reference systems were aligned, we would have $\alpha_{sp} = \alpha'_{sp}$. The rotation that would be needed to bring the array into alignment with the *world reference system* is (see Figure 2.9)

$$\text{rot} = \alpha_{sp} - \alpha'_{sp}. \quad (2.24)$$

Conversely, any estimated azimuth relative to the *array reference system* ($\hat{\alpha}$) may be converted to an azimuth relative to the *world reference system* ($\hat{\alpha}'$) by computing

$$\hat{\alpha}' = \hat{\alpha} - \text{rot}. \quad (2.25)$$

Since the positions of the arrays and the calibrating speaker are known (relative to the *world*), so is the azimuth of the speaker relative to each array in the *world reference system* (α'_{sp}). Therefore, by estimating α_{sp} using the sounds produced by the speaker, we immediately obtained an estimate for rot.

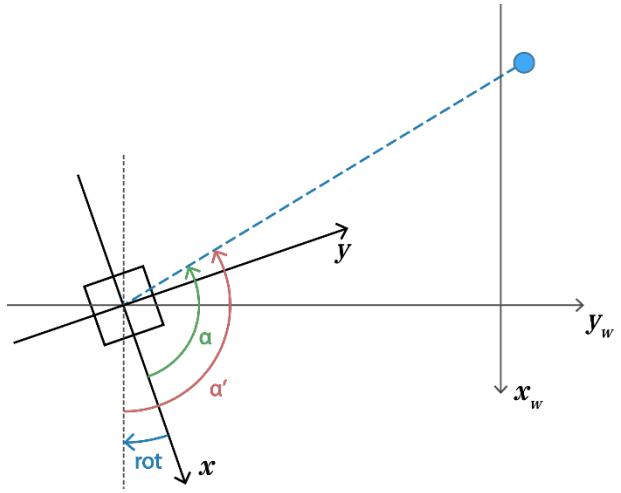

Figure 2.9 **Orientation of the array relative to the *world reference system* (x_w, y_w).** The calibration speaker is represented by the blue circle, and array *1* by the square on the left. The green arrow represents the azimuth α of the calibration speaker relative to the *array reference system*; the red arrow (α') represents the azimuth of the speaker relative to the array in the *world reference system*. rot is the rotation that would align the *array reference system* with the *world reference system*.

We used artificial sounds to estimate a credible region for the bearing of the speaker (relative to each *array reference system*) as described in 2.6.3.

We then used (2.22) and (2.23) to obtain the limits ($\alpha_{sp_{low}}$ and $\alpha_{sp_{up}}$) of the credible intervals for the azimuth (α_{sp}).

Finally, the limits of the credible intervals for rot were computed as

$$\mathrm{rot}_{low} = \alpha_{sp_{low}} - \alpha'_{sp} \qquad and \tag{2.26}$$

$$\mathrm{rot}_{up} = \alpha_{sp_{up}} - \alpha'_{sp} \quad . \tag{2.27}$$

2.7.3. Incorporating array orientation uncertainty in the azimuth estimate

Since the *credible interval of probability* p is an interval in which a given parameter has probability p of being included, it is possible to obtain intervals with a *minimum probability* for functions of those parameters as long as they are independent.

Suppose we used a set of artificial sounds to estimate a credible interval for the array orientation (rot), and then used biological data to estimate a credible interval for the azimuth α of a sound or group of sounds biological in origin. Let us define the events

$$\begin{aligned} A\text{:} \quad & \alpha \in [\alpha_{low}, \alpha_{up}]; \\ R\text{:} \quad & rot \in [rot_{low}, rot_{up}];\ and \\ A'\text{:} \quad & \alpha' \in [\alpha'_{low}, \alpha'_{up}]\,. \end{aligned}$$

Where $[\alpha_{low}, \alpha_{up}]$ is the credible interval of probability p_a for α; $[\mathrm{rot}_{low}, \mathrm{rot}_{up}]$ is the credible interval of probability p_r for the orientation of the array relative to the *world*;

$$\alpha'_{low} \equiv \alpha_{low} - \mathrm{rot}_{up} \quad and \tag{2.28}$$

$$\alpha'_{up} \equiv \alpha_{up} - \mathrm{rot}_{low} \quad . \tag{2.29}$$

All information about A and R originate either from the *a priori* probabilities for the bearings of each source, or from the data obtained from the biological sounds (for A) and from the artificial sounds (for R). Since all those sources of information are mutually independent, it follows that A and R are also independent and therefore

$$P(A \cap R) = P(A)P(R) = p_\alpha p_r \ . \tag{2.30}$$

α'_{low} and α'_{up} are defined in such a way ((2.28) and (2.29)) that $A' \supseteq A \cap R$. It follows that

$$P(A') \geq p_\alpha p_r \ . \tag{2.31}$$

We may therefore refer to $[\alpha'_{low}, \alpha'_{up}]$ as the *lower-bounded credible interval* of minimum probability $p_s = p_\alpha p_r$.

Note that the *lower-bounded credible interval* $[\alpha'_{low}, \alpha'_{up}]$ is a more conservative interval estimation than a credible interval, since the probability of the parameter (α') being included in it is higher (we just do not quantify how much higher) than any $p_s = p_\alpha p_r$ we choose to use.

2.7.4. Lower-bounded credible (x_w, y_w) regions

For each sound arriving at an array, if the position of the array relative to the world (x_a, y_a) is known, an interval of minimum probability for the azimuth relative to the world reference system $[\alpha'_{low}, \alpha'_{up}]$ has been determined, and the sound is assumed to not have been reflected or refracted laterally (it may have been reflected off the surface or bottom); then the event A' may also be defined (Figure 2.10):

> A' : The vertical projection of the sound source onto the $x0y$ plane is included in the area between the 2 rays originating at $(x_a, y_a, 0)$, with directions defined by α'_{low} and α'_{up}.

That is the same as saying the source of the sound is either above, below or on the area of the horizontal plane between the 2 rays originating at the array position with directions defined by α'_{low} and α'_{up}.

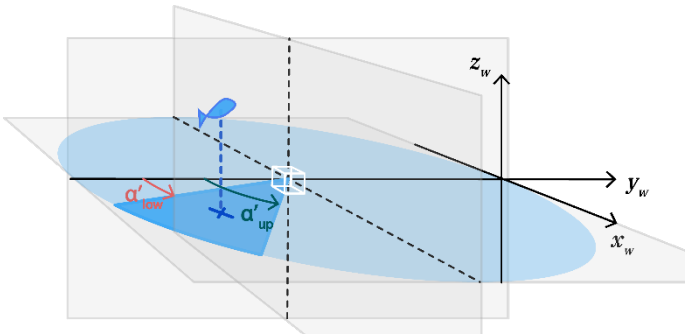

Figure 2.10 **Vertical projection of the sound source position onto the $x0y$ plane.** The vertical projection of the source position ontp the horizontal plane (marked with an x) is included in the area (darker blue) between the 2 rays originating at $(x_a, y_a, 0)$, with directions defined by α'_{low} and α'_{up}. The white cube represents the position of the array.

Furthermore, if a sound is detected by 2 different arrays, there will be 2 such areas (a_1 and a_2), one originating at each array position $(x_{a1}, y_{a1}, 0)$ and $(x_{a2}, y_{a2}, 0)$, and bound by the pairs of rays defined by $(\alpha'_{low_1}, \alpha'_{up_1})$ and $(\alpha'_{low_2}, \alpha'_{up_2})$. Provided p_s is high enough, these areas should usually intersect at an area of the horizontal plane (a_c). Let us then define three more events:

C: The vertical projection of the position of the sound source onto the $x_w 0 y_w$ plane is included in the area a_c;

A'_1: The vertical projection of the sound source onto the $x_w 0 y_w$ plane is included in the area between the 2 rays originating at $(x_{a1}, y_{a1}, 0)$, with directions defined by α'_{low_1} and α'_{up_1}; and

A'_2: The vertical projection of the sound source onto the $x_w 0 y_w$ plane is included in the area between the 2 rays originating at $(x_{a2}, y_{a2}, 0)$, with directions defined by α'_{low_2} and α'_{up_2}.

Then $C = A'_1 \cap A'_2$. Therefore, since $(\alpha'_{low_1}, \alpha'_{up_1})$ and $(\alpha'_{low_2}, \alpha'_{up_2})$ are obtained with independent priors and assuming the propagation uncertainties from the source to each array are independent,

$$P(C) = P(A'_1) \cdot P(A'_2) \geq p_s^2 \ . \tag{2.32}$$

In this work, we used $p_\alpha = p_r = 0.99$ which means $p_s^2 = 0.961$. The areas obtained are therefore areas which include (with probability higher than 0.961) the vertical projection of the position of the fish that produced a given sound or group of sounds.

To plot a_c, we used the `R` function `st_polygon()` from the package `sf` (Pebesma, 2018) to define 2 triangles with lateral sides defined by $(\alpha'_{low_1}, \alpha'_{up_1})$ and $(\alpha'_{low_2}, \alpha'_{up_2})$, apices at (x_{a1}, y_{a1}) and (x_{a2}, y_{a2}), and height large enough so the bases would not be included in the plotting area (specifically $100\ m$). The function `st_intersection()` of that same package was then used to obtain a_c (or at least the part of it that fit in the plot area), and the function `geom_sf()`, also from package `sf`, was used in conjunction with `ggplot()` from the `ggplot2` package (Wickham, 2016) to produce the plots presented in the results section.

3. Results

In section 3.1 we show several results obtained regarding groups of vocalisations assumed to have been produced by single stationary individuals. In 3.1.1, to check if the data is consistent with the assumption of normality, we show the distribution of the *observed TDOA*s conditional on $\hat{z}_{ki}$ (the estimated membership of each observation i to each group k estimated as described in 2.5.3). Then, in 3.1.2 we look at the correspondence between the groups obtained with the data from each array. If such groups can be interpreted as individual calling males, we expect them to be coherent across arrays. Finally, in 3.1.3 we show the *lower-bounded credible regions* estimated for each individual.

Section 3.2 describes the positions estimated for the origin of single vocalisations, without assuming they belong to any specific individual. Since these have been attributed to specific individuals in section 3.1, they are presented as belonging to such individuals. Yet, it should be possible to obtain this type of estimate even for vocalisations that could not be assigned to individual fish (including vocalisations made by fish such as the meagre which for which the assumption of stationarity is not necessarily valid). Therefore we organise section 3.2 based on the grouping from section 3.1 not because it depends on such grouping, but in order to make apparent whether they contradict or validate each other. In this section we also illustrate the effect of choosing a different lower bound for the credibility value (specifically a lower one) on the extension of the estimated regions.

3.1. Expectation-Maximization (E-M) algorithm

3.1.1. *TDOA* distributions

Since it is an underlying assumption of the method described here that the *TDOA*s follow a multivariate Normal distribution, we tested this hypothesis with the data collected.

Figure 3.1 and Figure 3.2 show the observed marginal distributions of each *TDOA* (coloured histogram), overlaid by the theoretical marginal distributions estimated by the E-M algorithm (black curves). The empirical distributions are represented as histograms weighted by the membership to each group k as estimated by the EM algorithm.

To test the hypothesis of normality within each group, each observation i was assigned to the group k for which $\hat{z}_{ki} = \max_k(\hat{z}_{ki})$. The samples defined by these groups were then used to run a *Henze-Zirkler test* (Henze and Zirkler, 1990) of multivariate normality. The *Bonferroni-corrected* p values of the test were non-significant (at $\alpha = 0.01$) for all groups in both arrays.

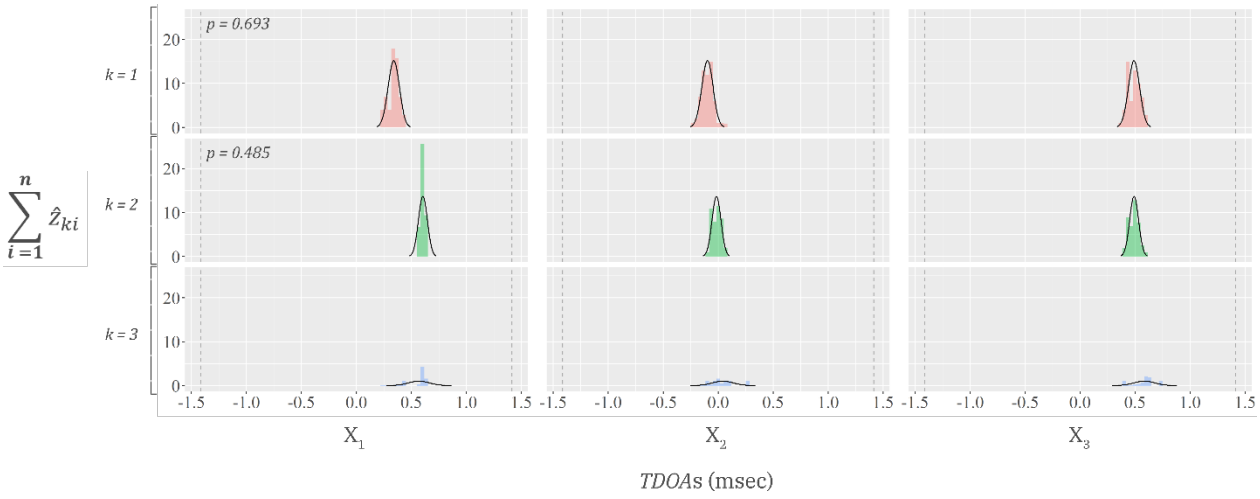

Figure 3.1 **Histogram of observed *TDOA*s from *array 1*, weighted by the estimated *membership* to each group k (putative individual).** The grey dashed vertical lines represent the limits of the possible range for μ_i given the distance between hydrophones $|r_i| = \sqrt{2}$ for all $i = (1, 2, 3)$, the estimated sound velocity $C = 1527$, and the model (2.3). The black curves are the densities of the estimated distributions, scaled so the total area is $\sum_{i=1}^{n} \hat{z}_{ki}$. Also represented on the upper left corner of each row, the p-value of the *Henze-Zirkler test (Henze and Zirkler, 1990)* for multivariate normality, *Bonferroni-corrected* for multiple testing. P-values are not shown for the group with the least number of observations, because it is not assumed to correspond to a single individual, and therefore it is not assumed to follow a single normal distribution.

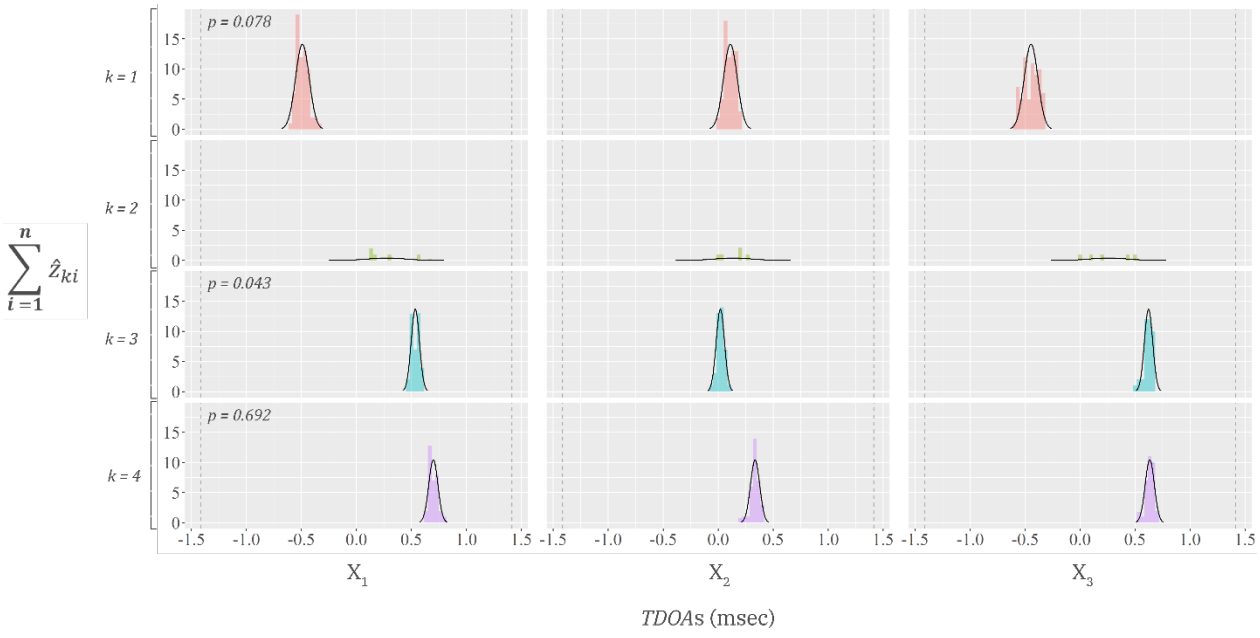

Figure 3.2 **Histogram of observed *TDOA*s from *array 2*, weighted by the estimated *membership* to each group k (putative individual).** The other components of the plot are as in *Figure 3.1*.

3.1.2. Classification concordance between arrays

Let us exclude from the following analysis the smallest group of sounds detected at each array. This exclusion is justified by assuming the group with fewer observations is aggregating sounds from multiple individuals which produced few vocalisations during the data collection period. Examining Figure 3.3b, it is apparent that to each of the 2 groups that emerged from *array 1* data, corresponds a specific group from

array 2, and that neither of these overlaps group 1 from *array 2*. Therefore, three distinct groups (A, B and C) emerge from crossing the data coming from both arrays.

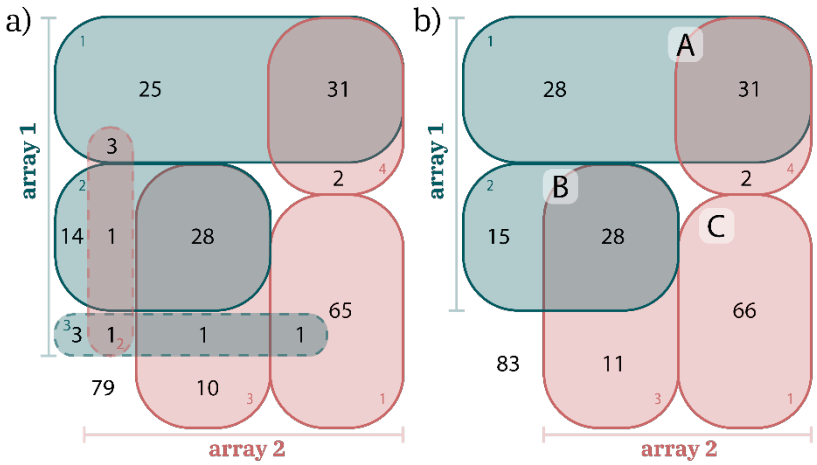

Figure 3.3 **Venn diagrams comparing the attribution of sounds to individuals between arrays.** a) includes all groups of sounds output by the E-M algorithm, while b) excludes those with the least observations for each array. Apart from the 2 smallest groups, which probably do not correspond to individual fish, the groups are consistent between arrays. The small coloured numbers represent the group number (*k*), the black numbers inside a single balloon represent the number of sounds for which it was possible to estimate a 3D bearing for one array only. The black numbers in the intersection of two balloons represent the number of sounds for which a 3D bearing was estimated relative to both arrays (these are the sounds for whose location it was possible to obtain a minimum credibility region). In panel *a* the black number outside the balloons represent the sounds present in the file for which it was not possible to obtain bearing estimates relative to any array, while in panel b) it represents those with no bearing estimates, plus those assigned only to the excluded groups. The letters A, B and C correspond to the groups that emerge from crossing the data from both arrays.

For 79 sounds the bearing was not obtained, while for an additional 119 sounds it was obtained only for a single array (see Figure 3.3a). This was due to the waveform (see Figure 2.4a-e) being distorted by noise or multiple overlapping sounds, making it impossible to select corresponding peaks.

3.1.3. Positions of putative individuals/nests

The *TDOA*s attributed to each individual toadfish (by assuming they call from a fixed position i.e. the nest), were used to estimate their azimuths relative to each array (as described in subsection 2.6.1). The estimated azimuths of each individual were then crossed to obtain an estimate of its position.

The lower-bounded credible regions for both the azimuths and the position of the putative individual producing each group of vocalisations A and B are depicted in Figure 3.4 (individual A on panel b and individual B on panel d). Each lower-bounded credible region for the position of the fish comprises a finite area that encompasses the position of the calling fish (with probability P >= 0.961).

Individual C appears to have produced enough vocalisations detected by array 2 to form a third group. Yet, not enough of these were detected by array 1 for a matching group to be formed from data from this array. Nevertheless, it was still possible to obtain a credible region for its azimuth relative to array 2. This result conveys some information about the putative individual position (albeit in the form of an infinite radial sector of the plane instead of a finite area as for A and B - Figure 3.4f).

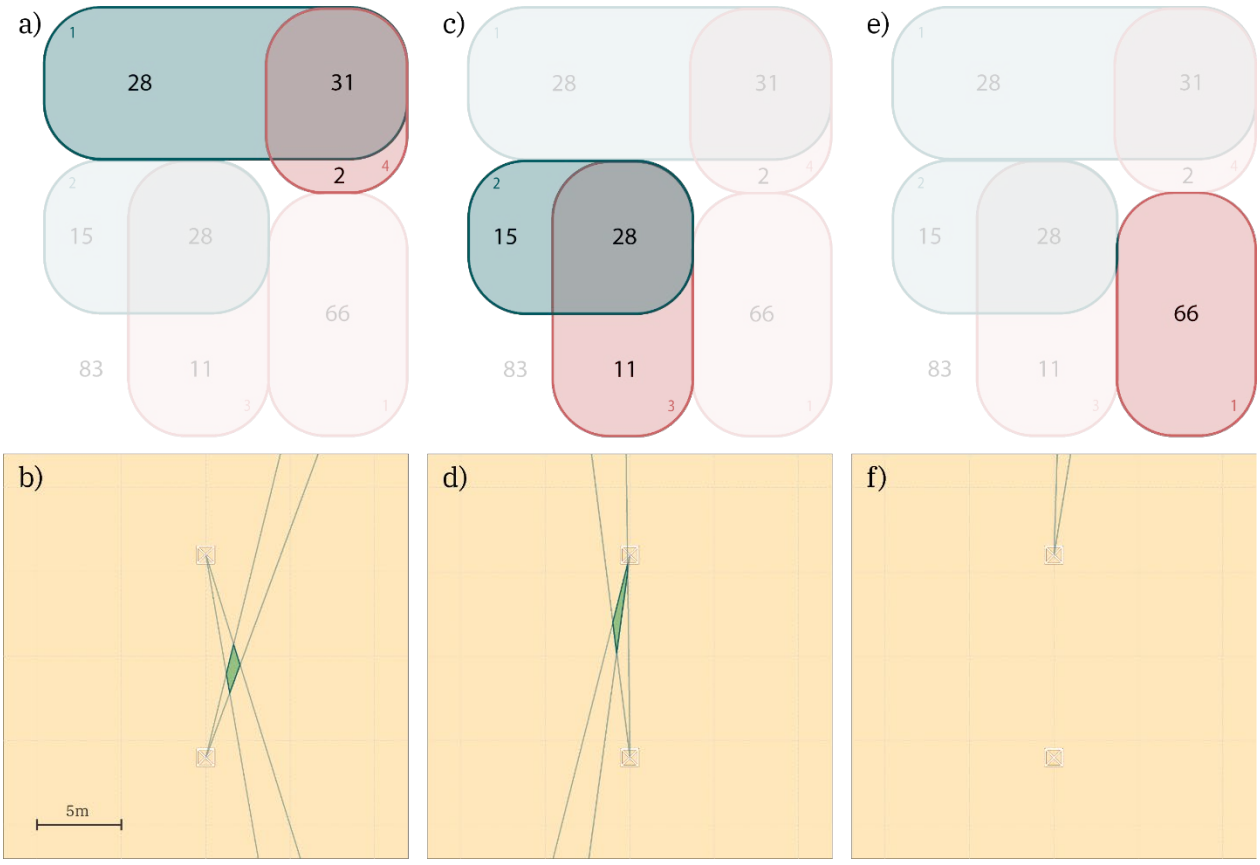

Figure 3.4 **Estimated area of origin of the sounds produced by each putative individual.** Panels a) and b) correspond to individual A, c) and d) to individual B, and e) and f) to individual C. a), c) and e) are the areas of the Venn diagram from *Figure 3.3* corresponding to each individual. b), d) and f) show the respective limits of the lower-bounded credible regions ($P >= 0.980$) for the azimuths (α') – green lines – and the lower-bounded credible regions with $P >= 0.961$ for the individual's position – green shaded area. Each array is represented by a white box drawn to scale. *Array 1* is the one closest to the lower margin of each panel, and *array 2* is closest to the upper margin.

3.2.Single sounds

In this section we look at the credible regions for the source position of individual sounds. To compare these results with those of section 3.1 as simply as possible, they are presented according to the grouping described above (A, B and C). Yet, note that it is possible to obtain this kind of results even for fish that produced a single vocalisation or that produced vocalisations while moving.

Figure 3.5 shows the waveforms, spectrograms, azimuth estimates and (when obtained) the position estimates of three vocalisations attributed to individual A. Panel a) represents one of the thirty-one vocalisations that had their bearing estimated relative to both arrays (see also Figure 3.4a), on panel b) is one of two that had their bearing estimated only relative to *array 2,* and on panel c) is one of 28 whose bearing was only estimated relative to *array 1.* First, we note the credible region for the sound source position (Figure 3.5a) is larger than, but includes, the region in Figure 3.4b. It is also apparent that the waveform has both a shape and an amplitude that is relatively uniform within each array (compare to Figure 3.6 and Figure 3.7). A greater proportion of red and yellow on the leftmost spectrograms reveals that the vocalisations of both individuals were more strongly received at *array 1* than at *array 2*. Both when the bearing could only be estimated for *array 1* (panel c) and when it could only be estimated for *array 2* (panel

b), the credible region for the azimuths is consistent with the credible regions of the sounds detected by both arrays (panel a).

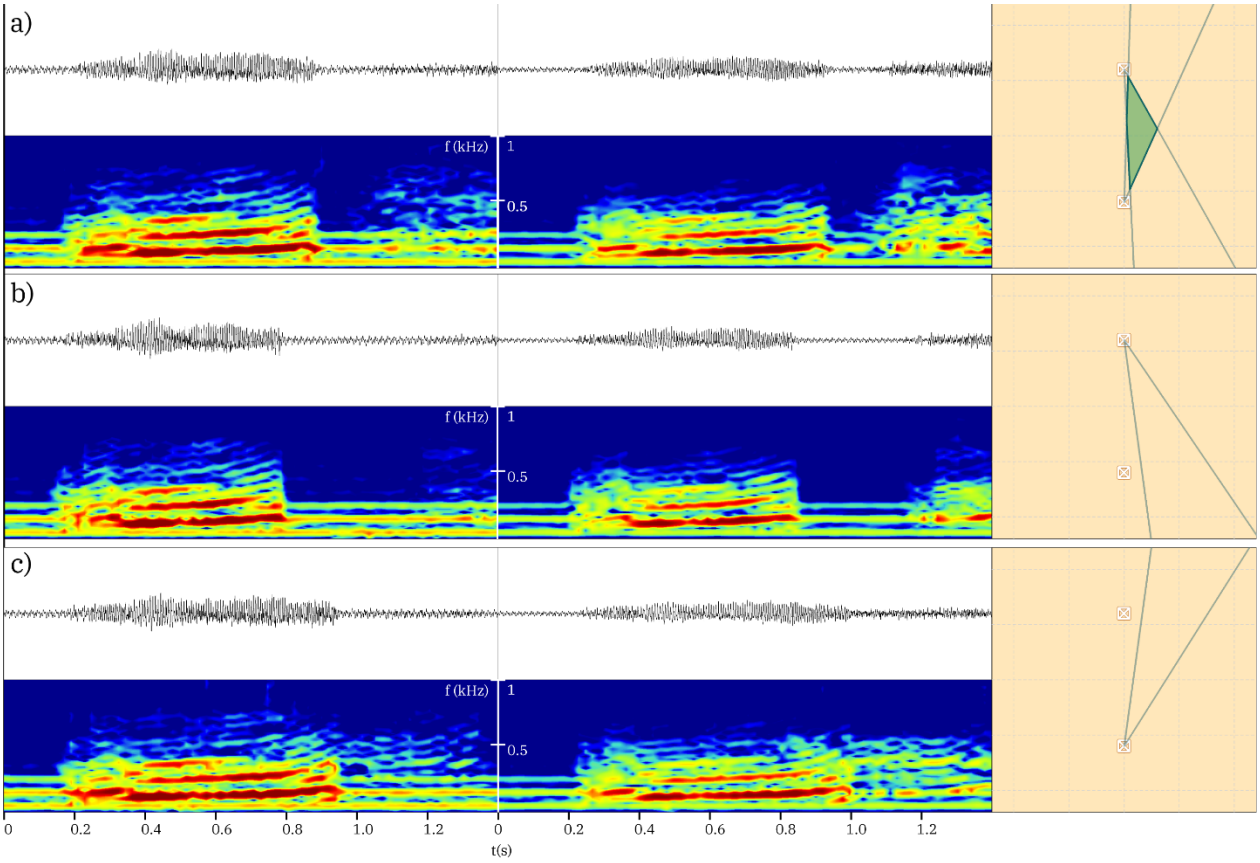

Figure 3.5 **Vocalisations attributed to putative individual A.** Each panel (a - e) corresponds to a vocalization. The vocalisation on panel a) has had its bearing estimated relative to both arrays, that on panel b) only relative to *array 2*, and that on panel c) only relative to *array* 1. On the left, the *waveform* and *spectrographic* views of the sound recorded at the reference hydrophone of *array 1*. At the centre, the same representations, as recorded by *array 2*. On the right, the limits of the lower-bounded credible regions ($P >= 0.980$) for the azimuths (α') of the location where the sound was produced relative to each array are represented as green lines. Also on the right, the lower-bounded credible region ($P >= 0.961$) for the position where the sound was produced is depicted as a green area. The spectrograms were computed (using the `seewave` R package) with a 4096 points *Hanning* window, padded with 2048 zeroes on each side, and 50% overlap. In the spectrograms, warmer colours correspond to higher sound pressure levels.

Very similar observations can be made about individual B (Figure 3.6): the estimated region for single vocalisations is larger but includes the region estimated using the group of vocalisations (Figure 3.4d); the shape and amplitude of the waveforms are similar within each array; the signal is stronger on *array 1* than on *array 2*; and the azimuth estimates when obtained for a single array are consistent with the position estimates when azimuths to both arrays are obtained.

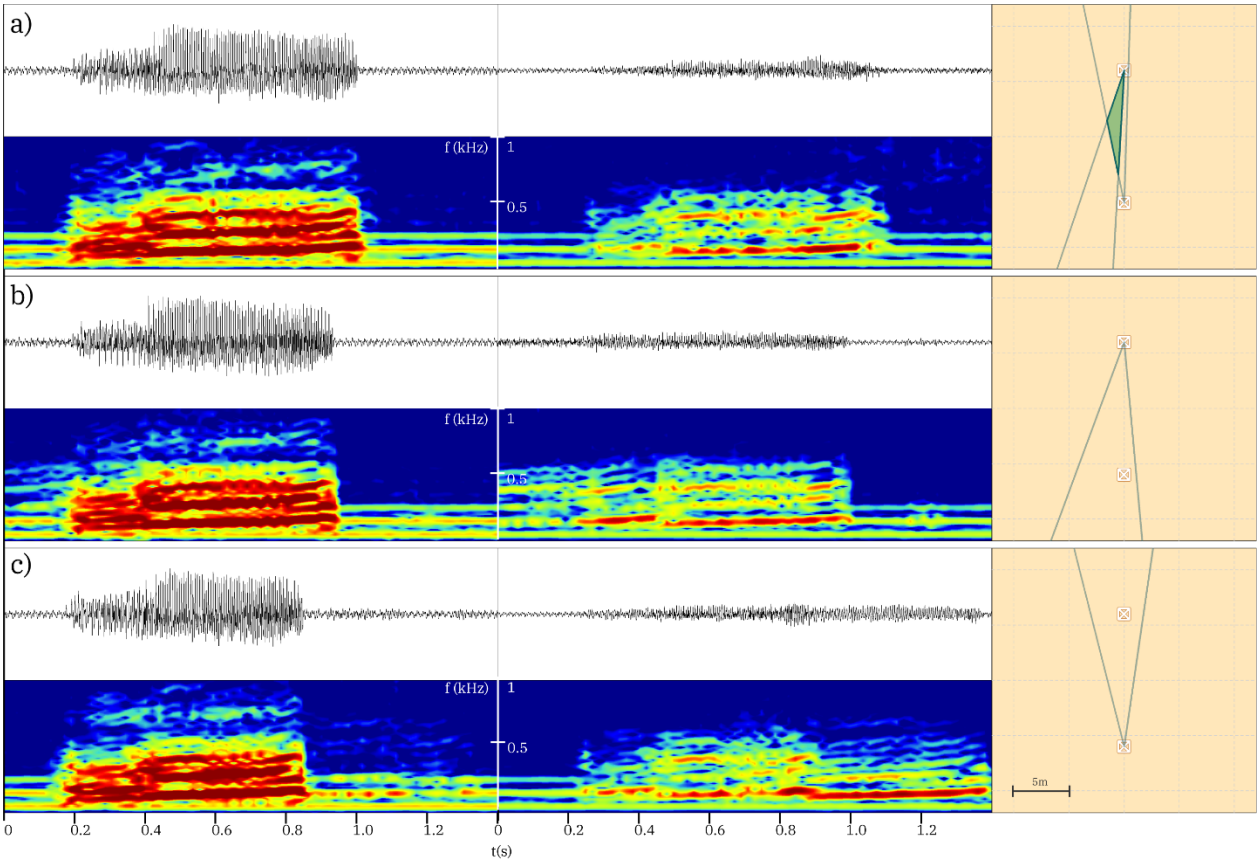

Figure 3.6 **Vocalisations attributed to putative individual B.** Each panel (a - e) corresponds to a vocalization. The vocalisation on panel a) has had its bearing estimated relative to both arrays, that on panel b) only relative to *array 2*, and that on panel c) only relative to *array* 1. On the left, the *waveform* and *spectrographic* views of the sound recorded at the reference hydrophone of *array 1*. At the centre, the same representations, as recorded by *array 2*. On the right, the limits of the lower-bounded credible regions ($P >= 0.980$) for the azimuths (α') of the location where the sound was produced relative to each array are represented as green lines. Also on the right, the lower-bounded credible region ($P >= 0.961$) for the position where the sound was produced is depicted as a green area. The spectrograms were computed (using the `seewave` R package) with a 4096 points *Hanning* window, padded with 2048 zeroes on each side, and 50% overlap. In the spectrograms, warmer colours correspond to higher sound pressure levels.

The vocalisations attributed to individual C (Figure 3.7) have a much lower SNR on *array 1* than those attributed to A and B, as can be seen in both the spectrogram and waveform representations of the sound recorded on *array 1*. The estimated azimuth of individual C relative to *array 2* is opposite to the direction of *array 1*. This is consistent both with the azimuth estimate for individual C (Figure 3.4f) and with the fact that it is almost not detected by *array 1*.

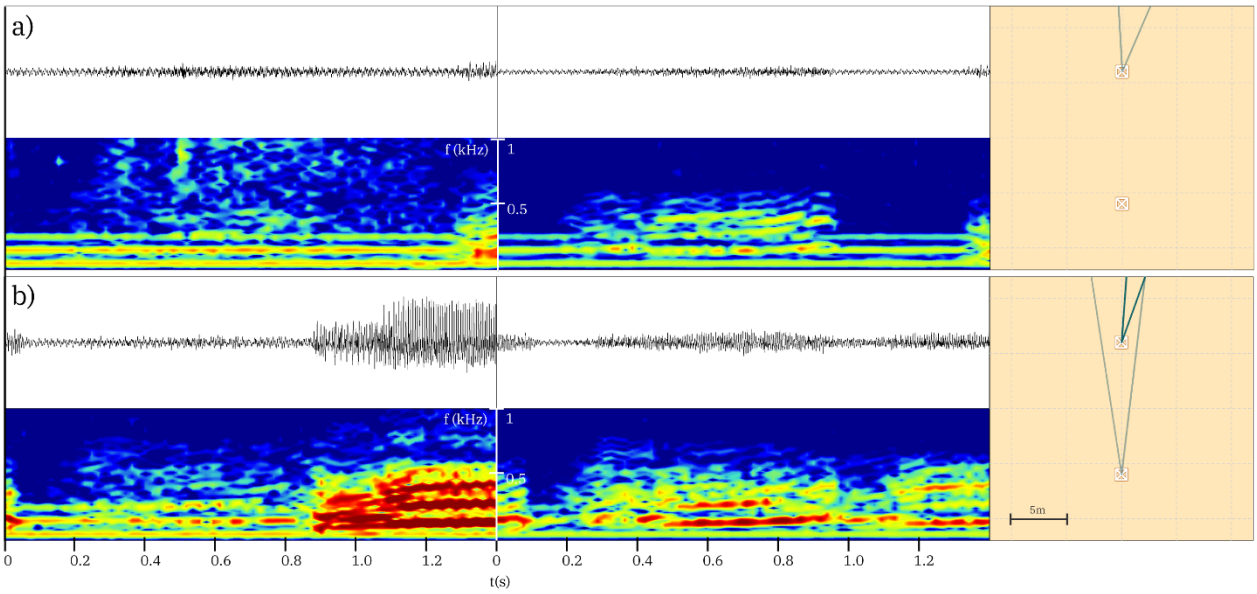

Figure 3.7 **Vocalisations attributed to putative individual C.** Each panel (a and b) corresponds to a vocalization. The vocalisation on panel a) has had its bearing estimated relative only to *array 2*, while that on panel b) had it estimated for both arrays. the vocalisation represented in b) was not attributed to individual B for *array 1*, probably because it was the only vocalisation from individual B to have its bearing estimated relative to that array. On the left, the *waveform* and *spectrographic* views of the sound recorded at the reference hydrophone of *array 1*. At the centre, the same representations, as recorded by *array 2*. On the right, the limits of the lower-bounded credible regions ($P >= 0.980$) for the azimuths (α') of the location where the sound was produced relative to each array are represented as green lines. Also on the right, the lower-bounded credible region ($P >= 0.961$) for the position where the sound was produced is depicted as a green area. The spectrograms were computed (using the `seewave` R package) with a 4096 points *Hanning* window, padded with 2048 zeroes on each side, and 50% overlap. In the spectrograms, warmer colours correspond to higher sound pressure levels.

The smallest groups of vocalisations detected at each array were not attributed to any individual since they were assumed to contain vocalisations from possibly multiple individuals that each produced too few vocalisations to form a group. One example is represented in Figure 3.7b). This is one of three vocalisations that appears to have been produced by individual C, for which it was possible to obtain the bearing relative to *array 1* as well as *array 2.* Three further examples are shown in Figure 3.8 that appear to have different explanations. All three vocalisations in this figure were attributed to individual A only from the data gathered by array 1, while they were part of the excluded group on *array 2* data. Judging from the position estimates, the vocalisation represented on panel c) appears to have been produced by individual A, while those represented on panels a) and b) were probably produced by a fourth individual aligned with individual A with respect to *array 1*, but not with respect to *array 2*.

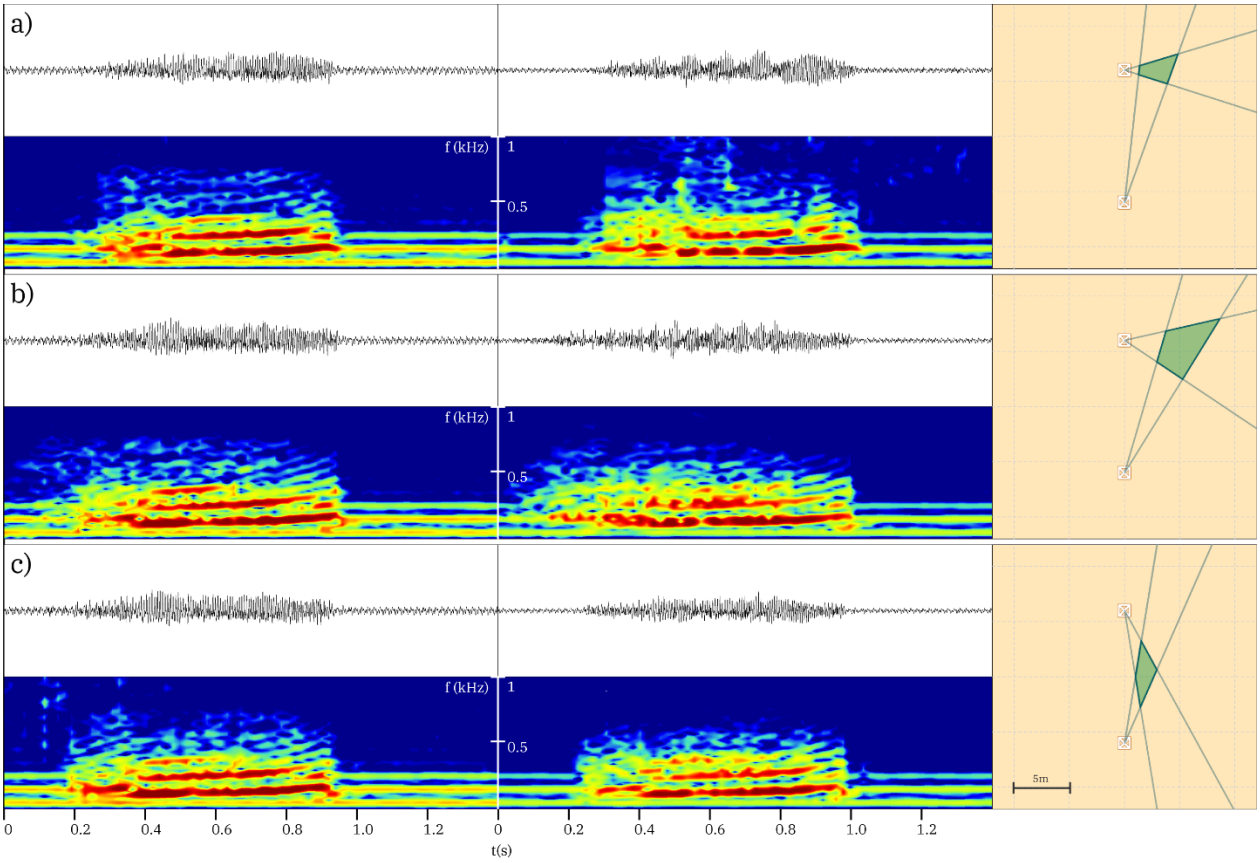

Figure 3.8 **The three vocalisations detected by both arrays and assigned to *group 1* on *array 1* (individual A) and *group 2* on *array 2* (excluded group).** On the left, the *waveform* and *spectrographic* views of the sound recorded at the reference hydrophone of *array 1*. At the centre, the same representations, as recorded by *array 2*. On the right, the limits of the lower-bounded credible regions ($P >= 0.980$) for the azimuths (α') of the location where the sound was produced relative to each array are represented as green lines. Also on the right, the lower-bounded credible region ($P >= 0.961$) for the position where the sound was produced is depicted as a green area. The spectrograms were computed (using the `seewave` R package) with a 4096 points *Hanning* window, padded with 2048 zeroes on each side, and 50% overlap. In the spectrograms, warmer colours correspond to higher sound pressure levels.

Even though 0.961 was the lower-bounded credible value chosen for the estimates chosen in this work, this value is arbitrary, and so it is possible to obtain regions with lower credibility and smaller area. As an example, Figure 3.9 compares the regions obtained for two vocalisations with two different minimum credibilities ($P \geq 0.961$ and $P \geq 0.80$).

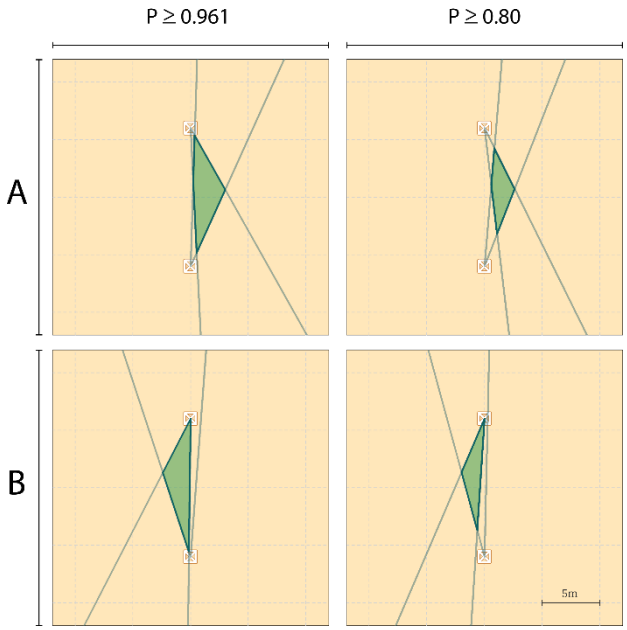

Figure 3.9 **Lower-bounded credible regions with ($P \geq 0.961$ and $P \geq 0.80$) of two vocalisations produced by different individuals.** The two top panels correspond to a vocalisation produced individual A, and the bottom two to a vocalisation produced by individual B.

4. Discussion

With this work, we were able to develop and deploy a system capable of estimating the position of vocalising fish, as well as intuitively conveying the uncertainty associated to such an estimate. The results presented in the previous chapter are promising in that they are consistent with what we would expect from previous knowledge of the fish location spots. In particular, the fact that it was possible to discriminate groups of sounds coming from different directions (3.1.1) and that those groups were consistent between arrays, even though they were estimated independently from the recordings obtained at each array (3.1.2), validates the interpretation of the groups as sounds produced by specific individuals (or at least produced by individuals at a specific location). According to this interpretation different vocalizations could be attributed to three different individuals, A, B and C, even though few vocalisations made by individual C could be used to estimate a bearing from *array 1*. This would have resulted in no group emerging from *array 1* data, corresponding to individual C and it is consistent with the estimated azimuth of individual C relative to *array 2* (opposite the direction of *array 1*). It is also consistent with almost no bearing estimates relative to *array 1* having been obtained for sounds produced by this individual. Amorim and Vasconcelos (2008) suggest that within a short time period individual recognition based on acoustic cues is possible, as intra-individual variation of boatwhistle acoustic features is much smaller than that between different subjects. This too is consistent with the results obtained, since both the waveforms and the spectrograms of the vocalisations (which visually reflect those acoustic features) are apparently more similar within each group than between groups (3.1.3). Finally, the fact that the location estimated for each single vocalisation (section 3.2) is consistent with (although less precise than) the estimate for the corresponding groups is in in line with multiple males vocalising from fixed positions near the arrays.

The smallest group of vocalisations detected by each array was, as described above, discarded in the sense that it was not assumed to correspond to a single individual. Yet it is of note that, once this assumption is discarded, the estimated locations of the single vocalisations attributed to those groups are compatible with what we would expect to observe. Indeed, in almost all cases in which a vocalisation is attributed to one of these groups in one array and to a larger group in the other array, both the estimated location and the acoustic features (as evaluated visually from the waveform and spectrogram) are consistent with the location of the individual corresponding to the larger group. The only exception is consistent with having been produced by a fourth individual aligned with A with respect to *array 1* but not with respect to *array 2*. That being the case, if more vocalisations by this individual had been produced, an extra group would have emerged from array 2 data.

Since studies using passive acoustic localization for fish triangulation are still scarce, and the methods employed are constrained by their specific features, it is hard to directly compare our method to the existing ones. Yet, some features in which it differs from previous work regarding fish localization by passive acoustic methods include (1) the usage of multiple arrays to overcome the limitation of point localization to distances of the same order of magnitude as the distance between the hydrophones of the array, (2) the output of a readily interpretable result that informs on both the estimated location and the associated uncertainty, (3) the modelling of the dependency structures between observed *TDOA*s of different pairs of hydrophones, and (4) the estimation of *observed TOA* variance using fish known to vocalise from fixed positions.

In what concerns the usage of multiple arrays, to our knowledge no other published work on fish localization using passive acoustics has employed such an approach. This is not surprising as the choice of this approach

stemmed from logistical limitations that were specific to the problem that motivated the development of this specific system. Namely, we required a system that could be deployed from a pier and would be able to locate meagre located a few tens of meters away, using 2 non-synchronised recorders capable of recording 4 simultaneous channels each. In other studies that use passive acoustics data to localise fish, either the hydrophones were deployed over a large area (e.g. Locascio and Mann, 2011; Parsons et al., 2009) when fish were distant, or used small arrays when the vocalising fish were close by. In this case the distance between hydrophones in the array was not much smaller than the distance between the hydrophones and the fish (e.g. Mouy et al., 2018; Putland et al., 2018). In contrast, our methodology allows the estimation of fish positions using independent recording systems connected to arrays whose dimensions are much smaller than their distance from the vocalising fish. In other words, it has the advantage of using a cost-effective small set of arrays with relatively low logistic requirements, while offering the advantage of larger arrays to locate fish at greater distances.

As for producing readily interpretable outputs including information regarding the uncertainty (specifically interval/region estimates), we also could not find any references in the literature regarding fish localization. At least one work on cetacean localization using passive acoustics does produce interval estimates for the parameters describing the location of the animal (Nosal and Frazer, 2007). However, the interpretation of such intervals is not at all straight-forward. Specifically, in this work Nosal and Frazer produce confidence intervals for each parameter *x, y* and *z*, from a grid-based approximation of the conditional likelihood of those parameters (conditional on the maximum likelihood estimate of the other two, and a best-guess for the *observed TOA variance*). In contrast, the method presented in this dissertation outputs *a region in which the vocalising animal is, with probability higher than a given predetermined value*. Furthermore, this value is arbitrary, meaning that the extension of that region can be reduced indefinitely, as long as one is willing to reduce the probability of it actually including the position of the fish.

The way in which the *observed TDOA* dependency structure is modelled varies across the scarce existing literature. In some cases it is not clearly specified (e.g. Locascio and Mann, 2011), in other papers, such as Parsons et al. (2010) and Putland et al. (2018), independent *observed TOA*s are implicitly assumed, while in a few publications (Gervaise et al., 2019; Mouy et al., 2018) the assumption of independent *observed TDOA*s is made explicit. Furthermore, among those who assume independent *observed TOA*s, there are different assumptions regarding their variance and consequently different approaches to its estimation and to the characterisation of the localization error. Putland et al. (2018), for instance, apparently assume that the *observed* TOAs differ less than 0.001s from the expected *TOA*s given the true location of the source. They compute the expected *TOA*s if the source was $5\ m$ from the array in each direction (0 to 360°, with 1° steps), simulate 1000 *TOA*s for each of those directions with the "*uncertainty* [...] *taken as* $\pm 0.001s$*"* and use the *mean Euclidean distance* between the "*actual"* (i.e. simulated) position and the *"calculated"* (i.e. estimated from the simulated *TOA*s) position to describe the uncertainty in the fish location estimate. It is not clear what distribution is used to estimate the *TOA*s, or to what parameter the 0.001s "*uncertainty"* refers to. Parsons et al. (2010), on the other hand, estimate standard errors for each *TOA* based on signal-to-noise ratios, signal overlap with other calls, and hydrophone position, but do not explain how. In the work presented in this text, we assume independent *TOA*s, and use sounds produced by individual male toadfish, which we assume to be stationary, to estimate their variance. We then use a worst-case guess, based on the variances estimated for the *TOA*s of the vocalisations of each individual toadfish, to produce *lower-bounded credible intervals* for the position of each single vocalisation. Such an approach allows the obtention of

region estimates even for animals which we cannot assume stationary (provided we assume the variance of the observed *TDOA*s of their vocalisations is not higher than that of the stationary toadfish).

4.1. Limitations and future prospects

Several limitations of the method presented in this work remained unaddressed, however, and will be the target of our work in the near future. One of these relates to the uncertainty in array orientation (section 2.7.2). In the deployment described in the present work, it was a reasonable assumption that the bases of the arrays were perfectly horizontal, which is to say, the direction of z (section 2.4) and that of z_w (section 2.4.2) were the same. This is true only because careful manual positioning of the arrays on the bottom was possible during low tide. Yet, it would be useful to deploy the system in locations where the bottom is not accessible, and its exact configuration is not well known. A solution might be to use an inclinometer attached to each array to measure the direction of gravity. Redefining the array reference system to have z coincident with the direction of gravity would ensure it is horizontal. A second limitation is the reliance on manual measurement of the *TDOA*s. This is both time consuming and arguably hinders reproducibility of results. An alternative would be the use of the cross-correlation function to automatically measure the *TDOA*s. This would require some adaptation of the model as the correlation structure of the observed *TDOA*s would no longer be as described in section 2.5.2. The propagation would still be correlated but the measurement of each *TDOA* would theoretically be independent. This means both σ^2_{prop} and σ^2_{meas} would have to be estimated separately.

Low precision (patent in the extension of the credible regions) is another problem addressable in several ways. Larger arrays would make the uncertainty on *TDOA* measurements be reduced in how much it propagates to uncertainty in bearing; high quality sensors might reduce the uncertainty in *TDOA* measurements by providing a higher signal-to-noise ratio; and an extra array, noncolinear with the other 2, could provide an extra lower-bounded credible region to intersect with the other two potentially decreasing the area of the intersection, namely when the source is near colinear with the arrays.

5. Concluding remarks

This dissertation was done within the context of a project with a longer temporal span, which aims at assessing and describing the effects of anthropogenic noise on the behaviour and reproductive fitness of the meagre and the toadfish in the Tagus estuary. Therefore, the results presented here represent a fundamental intermediate step towards the obtention and analysis of relevant biological data. Although, as noted above, some of the issues identified still need to be addressed, we are confident it will soon be possible to use this system to increase our knowledge of the problem that motivated its development.

The field of passive acoustic localization (especially in context of fish ecology and behaviour studies) is only in its infancy, and has the potential to bloom in the future. We hope to be able continue working on this topic. Although this thesis could only address methodological problems (leaving exciting biological questions to be tackled further down the road), it originated some answers as well as new questions. We look forward to finding the answers to these questions, and can already anticipate that new ones will arise. The only certainty is that we will not be bored.

6. References

Amorim, M.C.P., Conti, C., Sousa-Santos, C., Novais, B., Gouveia, M.D., Vicente, J.R., Modesto, T., Gonçalves, A., Fonseca, P.J., 2016. Reproductive success in the Lusitanian toadfish: Influence of calling activity, male quality and experimental design. Physiology & Behavior 155, 17–24. https://doi.org/10.1016/j.physbeh.2015.11.033

Amorim, M.C.P., Simões, J.M., Fonseca, P.J., 2008. Acoustic communication in the Lusitanian toadfish, *Halobatrachus didactylus* : evidence for an unusual large vocal repertoire. J. Mar. Biol. Ass. 88, 1069–1073. https://doi.org/10.1017/S0025315408001677

Amorim, M.C.P., Vasconcelos, R.O., 2008. Variability in the mating calls of the Lusitanian toadfish *Halobatrachus didactylus* : cues for potential individual recognition. Journal of Fish Biology 73, 1267–1283. https://doi.org/10.1111/j.1095-8649.2008.01974.x

Au, W.W., Hastings, M.C., 2008. Principles of marine bioacoustics. Springer.

Bénard, F., Glotin, H., Giraudet, P., 2011. Highly Defined Whale Group Tracking by Passive Acoustic Stochastic Matched Filter, in: Strumillo, P. (Ed.), Advances in Sound Localization. InTech. https://doi.org/10.5772/16173

Bolgan, M., Crucianelli, A., Mylonas, C., Henry, S., Falguière, J., Parmentier, E., 2020. Calling activity and calls' temporal features inform about fish reproductive condition and spawning in three cultured Sciaenidae species. Aquaculture 524, 735243.

Brüel&Kjær, 2019. Microphone Handbook, Technical Documentation. Brüel & Kjær, Naerum, Denmark.

Clark, C.W., Charif, R., Mitchell, S., Colby, J., 1996. Distribution and behavior of the bowhead whale, Balaena mysticetus, based on analysis of acoustic data collected during the 1993 spring migration off Point Barrow, Alaska. Report-International Whaling Commission 46, 541–554.

Crossin, G.T., Heupel, M.R., Holbrook, C.M., Hussey, N.E., Lowerre-Barbieri, S.K., Nguyen, V.M., Raby, G.D., Cooke, S.J., 2017. Acoustic telemetry and fisheries management. Ecol Appl 27, 1031–1049. https://doi.org/10.1002/eap.1533

De Robertis, A., Handegard, N.O., 2013. Fish avoidance of research vessels and the efficacy of noise-reduced vessels: a review. ICES Journal of Marine Science 70, 34–45. https://doi.org/10.1093/icesjms/fss155

Dos Santos, M.E., Modesto, T., Matos, R.J., Grober, M.S., Oliveira, R.F., Canário, A., 2000. Sound productoin by the Lusitanian Toadfish, *Halobatrachus didactylus*. Bioacoustics 10, 309–321. https://doi.org/10.1080/09524622.2000.9753440

Feistel, R., 2008. A Gibbs function for seawater thermodynamics for −6 to 80°C and salinity up to 120gkg–1. Deep Sea Research Part I: Oceanographic Research Papers 55, 1639–1671. https://doi.org/10.1016/j.dsr.2008.07.004

Friedman, J., Hastie, T., Tibshirani, R., others, 2001. The elements of statistical learning. Springer series in statistics New York.

Geoportal [WWW Document], n.d. . CoastNet. URL https://coastnet.pt// (accessed 9.9.20).

Gervaise, C., Lossent, J., Valentini-Poirier, C.A., Boissery, P., Noel, C., Di Iorio, L., 2019. Three-dimensional mapping of the benthic invertebrates biophony with a compact four-hydrophones array. Applied Acoustics 148, 175–193. https://doi.org/10.1016/j.apacoust.2018.12.025

Gibb, R., Browning, E., Glover-Kapfer, P., Jones, K.E., 2019. Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. Methods Ecol Evol 10, 169–185. https://doi.org/10.1111/2041-210X.13101

Giraudet, P., Glotin, H., 2006. Real-time 3D tracking of whales by echo-robust precise TDOA estimates with a widely-spaced hydrophone array. Applied Acoustics 67, 1106–1117. https://doi.org/10.1016/j.apacoust.2006.05.003

Glotin, H., Caudal, F., Giraudet, P., 2008. Whale cocktail party: Real-time multiple tracking and signal analyses. Canadian Acoustics 36.

Hartigan, J.A., Wong, M.A., 1979. Algorithm AS 136: A K-Means Clustering Algorithm. Applied Statistics 28, 100. https://doi.org/10.2307/2346830

Hawkins, A.D., Popper, A.N., 2018. Effects of Man-Made Sound on Fishes, in: Effects of Anthropogenic Noise on Animals. Springer Science+Business Media, New York, NY, pp. 145–178.
Hazen, E.L., Horne, J.K., 2004. Comparing the modelled and measured target-strength variability of walleye pollock, Theragra chalcogramma. ICES Journal of Marine Science 61, 363–377. https://doi.org/10.1016/j.icesjms.2004.01.005
Helble, T.A., Ierley, G.R., D'Spain, G.L., Martin, S.W., 2015. Automated acoustic localization and call association for vocalizing humpback whales on the Navy's Pacific Missile Range Facility. The Journal of the Acoustical Society of America 137, 11–21. https://doi.org/10.1121/1.4904505
Henze, N., Zirkler, B., 1990. A class of invariant consistent tests for multivariate normality. Communications in Statistics - Theory and Methods 19, 3595–3617. https://doi.org/10.1080/03610929008830400
Lagardère, J.P., Mariani, A., 2006. Spawning sounds in meagre Argyrosomus regius recorded in the Gironde estuary, France. J Fish Biology 69, 1697–1708. https://doi.org/10.1111/j.1095-8649.2006.01237.x
Locascio, J.V., Mann, D.A., 2011. Localization and source level estimates of black drum (Pogonias cromis) calls. The Journal of the Acoustical Society of America 130, 1868–1879.
Mann, D.A., Higgs, D.M., Tavolga, W.N., Souza, M.J., Popper, A.N., 2001. Ultrasound detection by clupeiform fishes. The Journal of the Acoustical Society of America 109, 3048–3054.
Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Moretti, D.J., Harris, D., Tyack, P.L., 2013. Estimating animal population density using passive acoustics. Biol Rev 88, 287–309. https://doi.org/10.1111/brv.12001
McElreath, R., 2020. Statistical rethinking: A Bayesian course with examples in R and Stan, 2nd ed. CRC press.
Misund, O.A., 1997. Underwater acoustics in marine fisheries and fisheries research. Reviews in Fish Biology and Fisheries 7, 1–34.
Mouy, X., Rountree, R., Juanes, F., Dosso, S.E., 2018. Cataloging fish sounds in the wild using combined acoustic and video recordings. The Journal of the Acoustical Society of America 143, EL333–EL339. https://doi.org/10.1121/1.5037359
Murteira, B., Antunes, M., 2012. Probabilidades e Estatística, 2nd ed. Escolar Editora, Lisboa.
Nakken, O., Olsen, K., 1977. Target strength measurements of fish. Ices.
Nosal, E.-M., Frazer, L.N., 2007. Sperm whale three-dimensional track, swim orientation, beam pattern, and click levels observed on bottom-mounted hydrophonesa). J. Acoust. Soc. Am. 122, 11.
Oppenheim, A.V., Schafer, R.W., Buck, J.R., 1999. Discrete-time signal processing, 2nd ed. ed. Prentice Hall, Upper Saddle River, N.J.
Parsons, M.J., McCauley, R.D., Mackie, M.C., Duncan, A.J., 2010. A Comparison of techniques for ranging close-proximity mulloway (Argyrosomus japonicus) calls with a single hydrophone. Acoustics Australia 38.
Parsons, M.J., McCauley, R.D., Mackie, M.C., Siwabessy, P., Duncan, A.J., 2009. Localization of individual mulloway (Argyrosomus japonicus) within a spawning aggregation and their behaviour throughout a diel spawning period 8.
Pebesma, E., 2018. Simple Features for R: Standardized Support for Spatial Vector Data. The R Journal 10, 439–446. https://doi.org/10.32614/RJ-2018-009
Pijanowski, B.C., Villanueva-Rivera, L.J., Dumyahn, S.L., Farina, A., Krause, B.L., Napoletano, B.M., Gage, S.H., Pieretti, N., 2011. Soundscape Ecology: The Science of Sound in the Landscape. BioScience 61, 203–216. https://doi.org/10.1525/bio.2011.61.3.6
Putland, R.L., Mackiewicz, A.G., Mensinger, A.F., 2018. Localizing individual soniferous fish using passive acoustic monitoring. Ecological Informatics 48, 60–68. https://doi.org/10.1016/j.ecoinf.2018.08.004
Radford, A.N., Kerridge, E., Simpson, S.D., 2014. Acoustic communication in a noisy world: can fish compete with anthropogenic noise? Behavioral Ecology 25, 1022–1030. https://doi.org/10.1093/beheco/aru029
Randall, R.B., 1987. Frequency analysis, 3. ed., rev.1. print. ed. Brüel & Kjaer, Naerum.

Roy, N., Simard, Y., Gervaise, C., 2010. 3D tracking of foraging belugas from their clicks: Experiment from a coastal hydrophone array. Applied Acoustics 71, 1050–1056. https://doi.org/10.1016/j.apacoust.2010.05.008

Slabbekoorn, H., Bouton, N., van Opzeeland, I., Coers, A., ten Cate, C., Popper, A.N., 2010. A noisy spring: the impact of globally rising underwater sound levels on fish. Trends in Ecology & Evolution 25, 419–427. https://doi.org/10.1016/j.tree.2010.04.005

Soetaert, K., Petzoldt, T., 2020. marelac: Tools for Aquatic Sciences.

Wang, D., Garcia, H., Huang, W., Tran, D.D., Jain, A.D., Yi, D.H., Gong, Z., Jech, J.M., Godø, O.R., Makris, N.C., Ratilal, P., 2016. Vast assembly of vocal marine mammals from diverse species on fish spawning ground. Nature 531, 366–370. https://doi.org/10.1038/nature16960

Wickham, H., 2016. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.