Localization of sound-producing fish in a water-filled tank

Antonin Novak, Petr Cisar, Michel Bruneau, Pierrick Lotton, and Laurent Simon

Citation: The Journal of the Acoustical Society of America **146**, 4842 (2019); doi: 10.1121/1.5138607
View online: https://doi.org/10.1121/1.5138607
View Table of Contents: https://asa.scitation.org/toc/jas/146/6
Published by the Acoustical Society of America

Localization of sound-producing fish in a water-filled tank

Antonin Novak,[1,a)] Petr Cisar,[2] Michel Bruneau,[1] Pierrick Lotton,[1] and Laurent Simon[1]
[1]*Laboratoire d'Acoustique de l'Université du Mans (LAUM, UMR CNRS 6613), 72000 Le Mans, France*
[2]*Laboratory of Signal and Image Processing, Institute of Complex Systems, South Bohemian Research Center of Aquaculture and Biodiversity of Hydrocenoses, Faculty of Fisheries and Protection of Waters, University of South Bohemia in České Budějovice, Zámek 136, Nové Hrady 37333, Czech Republic*

(Received 18 March 2019; revised 21 June 2019; accepted 4 July 2019; published online 31 December 2019)

In this paper, the authors introduce an algorithm for locating sound-producing fish in a small rectangular tank that can be used, e.g., in behavioral bioacoustical studies to determine which fish in a group is sound-producing. The technique consists of locating a single sound source in the tank using signals gathered by four hydrophones placed in the tank together with a group of fish under study. The localization algorithm used in this paper is based on a ratio of two spectra ratios: the spectra ratio between the sound pressure measured by hydrophones at two locations and the spectra ratio between the theoretical Green's functions at the same locations. The results are compared to a localization based on image processing technique and with video recordings acquired synchronously with the acoustic recordings. © *2019 Acoustical Society of America*.
https://doi.org/10.1121/1.5138607

[PJG] Pages: 4842–4850

I. INTRODUCTION

It has been known for many years that a wide range of fish species are capable of producing sounds and that sounds are important to them as part of their social behavior (Hawkins, 1986). Since then, the behaviors associated with acoustic communications have been the object of many ethological studies (Bertucci *et al.*, 2010; Fay, 2008; Ladich, 2015). The issues in studying the link between fish behavior and acoustic communication arise from the difficulty of identifying the sound producer. Although many studies can be conducted under laboratory conditions with low noise level, great visibility, and high quality equipment (hydrophones and video recordings), it is not easy to determine exactly which fish in a group produces the sound (Zelick *et al.*, 1999).

The goal of the research project presented in this paper is to develop a source localization technique in a water-filled fish tank that allows us to locate a single sound-producing fish using signals gathered by a few hydrophones set in the tank with a group of fish under study, as shown in Fig. 1.

Acoustic source localization is an inverse problem for which several methods have been developed: time difference of arrival (TDOA) (Daku *et al.*, 1992), beamforming (Johnson and Dudgeon, 1993), MUSIC (Asano *et al.*, 2001), sparse recovery (Malioutov *et al.*, 2005), and acoustic time reversal (Fink and de Rosny, 2001; Fink *et al.*, 2004). Many of these methods assume that acoustic propagation occurs in a free field or in an anechoic environment; some of them require very accurate estimates of time delays; other use larger microphone arrays (Kim *et al.*, 2015) or lead to a focal spot size limited to half a wavelength (Fink and de Rosny, 2001). None of these methods is suitable for localization in a small rectangular water-filled fish tank.

The tank is a highly reverberant acoustic environment exhibiting very strong and narrow resonances (Akamatsu *et al.*, 2002; Novak *et al.*, 2018). To deal with highly reverberating environments such as a water-filled fish tank, the set of Green's functions consisting of all frequency responses between each point source in the tank and each position of the hydrophone must be known either experimentally or by analytical or numerical modeling (Chardon *et al.*, 2015; Dokmanić and Vetterli, 2012; Nowakowski *et al.*, 2017).

The fish localization algorithm proposed in this paper takes advantage of a recently developed analytical formulation of the Green's function of a water-filled fish tank (Novak *et al.*, 2018). More specifically, the algorithm is based on a ratio of two spectra ratios: the spectra ratio between the sound pressure measured by hydrophones at two locations and the spectra ratio between the theoretical Green's functions at the same locations. The latter can be calculated in advance in the form of a Green's functions dictionary, or an optimization algorithm, such as a genetic algorithm, can be used instead of the full dictionary method to accelerate the localization (Stefanoiu and Ionescu, 2003). Both methods are discussed in the paper and are tested on experimental data provided by sound-producing fish.

The group of fish under study is a group of six adult croaking gourami (*Trichopsis pumila*). This gourami species is selected for the impulsive sounds they produce (Daugherty and Marshall, 1976). To evaluate the effectiveness of the proposed method, the results of the acoustic localization are compared with a video recording, allowing a localization based on an image processing technique.

The paper is organized as follows. The analytical formulation of the Green's function of a water-filled fish tank developed in Novak *et al.* (2018) is briefly recalled in Sec. II. Then, Sec. III describes the measurement conditions under which the acoustic and video recordings are made. Finally, the localization algorithm is presented in Sec. IV,

[a)]Electronic mail: antonin.novak@univ-lemans.fr

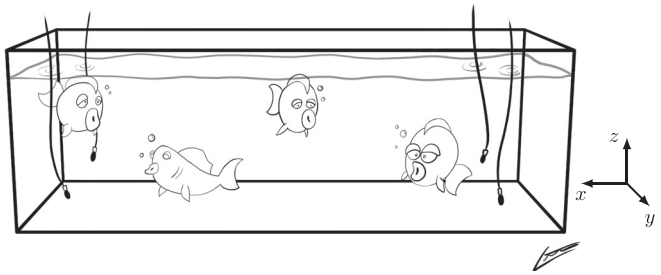


FIG. 1. A drawing of the water-filled tank with four hydrophones set in the tank together with a group of fish under study.

followed by a discussion in Sec. V. The robustness of the method is tested in the Appendix.

II. FISH TANK MODEL

The method of localization relies on the results of an accurate theoretical acoustic model of a liquid-filled tank (Novak *et al.*, 2018). This model expresses both the acoustic leakage through the walls (lossy and reacting walls) and the modal solutions for the sound pressure field and the acoustic velocity field. When a point source emits energy, the sound pressure response to a receiver position in a small liquid-filled tank is given by the product of the suitable source function and the Green's function, which is expressed here as an eigenfunction expansion.

A. Expression of Green's function

In the following, the x, y, and z axes are parallel to the orthogonal walls of the tank, the first two axes being in the horizontal plane and the last being vertical. The dimensions of the tank are noted L_x, L_y, and L_z, respectively, L_z being the water level of the tank. The thickness of the walls is noted h. Note that the upper surface of the liquid is a free surface that is not in contact with any wall. The superscripts and subscripts l, w, and a denote, respectively, the liquid, the wall, and the air, the subscript l (liquid) being removed when clearly not necessary. The parameters (ρ_l and c_l), (ρ_w and c_w), and (ρ_a and c_a) are the density and the speed of sound in the liquid, the wall, and the air, respectively. The angular frequency and the wavenumber in the water are noted, respectively, $\omega = 2\pi f$ and $k_l = \omega/c_l$.

The effects of the lossy and reacting vibrations of the walls on each acoustic mode in the brick of water (which include the energy leakage outside in the air) are specified by the small modal specific impedance-like $\zeta_{x_1 m}$ and $\zeta_{x_2 m}$ for the walls set at $x = 0$ (labeled x_1) and $x = L_x$ (labeled x_2), respectively (and similarly for the other walls and for the upper surface of the water):

$$\zeta_{x_1 m} = \frac{\xi_{m_x}^{(a,l)} + i\xi_{m_x}^{(w,l)} \tan(k_{m_x}^{(w)} h)}{1 + \xi_{m_x}^{(a,w)} \tan(k_{m_x}^{(w)} h)}, \tag{1}$$

with

$$\xi_{m_x}^{(\alpha,\beta)} = \frac{\rho_\alpha c_\alpha}{\rho_\beta c_\beta} \frac{k_{m_x}^{(\beta)}}{k_m^{(\beta)}} \frac{k_m^{(\alpha)}}{k_{m_x}^{(\alpha)}}, \tag{2}$$

where, given the eigen-angular frequency ω_m, the modal wavenumbers in the liquid, in the walls, and in the air are written as

$$k_m^{(l)} c_l = k_m^{(w)} c_w = k_m^{(a)} c_a = \omega_m, \tag{3}$$

and where, for any superscript α (l, w, or a), the components of the wavenumbers can be approximated by (Dirichlet conditions)

$$(k_{m_x}^{(\alpha)})^2 \simeq \left(\frac{\omega_m}{c_\alpha}\right)^2 - \left[\left(\frac{m_y \pi}{L_z}\right)^2 + \left(\frac{m_y \pi}{L_z}\right)^2\right], \tag{4}$$

(and similarly for the other components y and z) with m representing the triplet of indexes (m_x,m_y,m_z), $m = 0$ meaning m_x,m_y, $m_z = 0,0,0$.

The modal complex eigenvalues and the modal complex normalized eigenfunctions of the brick of water take, respectively, the following approximate form, to the lower order of the specific impedances $\zeta_{x_i m}$:

$$k_{m_x} L_x = m_x \pi + i(\zeta_{x_1 m} + \zeta_{x_2 m}), \tag{5}$$

(and similarly for the components y and z), and

$$\begin{aligned} \psi_m(x,y,z) = \sqrt{2^3/(L_x L_y L_z)} \sin\left(k_{m_x} x - i\zeta_{x_1 m}\right) \\ \times \sin\left(k_{m_y} y - i\zeta_{y_1 m}\right) \\ \times \sin\left(k_{m_z} z - i\zeta_{z_1 m}\right). \end{aligned} \tag{6}$$

Note that the effects of the walls labeled "2," those facing the walls labeled "1," are included in the components of the wavenumbers.

Finally, the Green's function for the pressure field $p_l(\vec{r}_0)$ at the receiver position $\vec{r}_0 = (x_0, y_0, z_0)$, corresponding to a harmonic (ω) point source located at $\vec{r}_s = (x_s, y_s, z_s)$ in the brick of water can be taken as ($k_l = \omega/c_l$):

$$G(\vec{r}_0, \vec{r}_s, \omega) = \sum_{m=0}^{\infty} \frac{\psi_m(x_s, y_s, z_s)}{k_m^2 - k^2} \psi_m(x_0, y_0, z_0). \tag{7}$$

B. Fish tank acoustics

The acoustic properties of the tank have a huge impact on the sound propagation in the tank. Figure 2 shows an example of three frequency response functions [calculated using the Green's function, Eq. (7)]. The position of the receiver (hydrophone) is located at the coordinates [$x_0 = 10$ cm, $y_0 = 10$ cm, $z_0 = 10$ cm] and three source positions are tested having the same y_s and z_s coordinates ($y_s = 10$ cm, $z_s = 10$ cm) and x_s coordinates being 20 cm (blue solid line), 30 cm (red dashed line), and 40 cm (green dashed-dotted line).

The frequency response functions depicted in Fig. 2 show that behavior at frequencies below the first resonance frequency (4.96 kHz) is very different from that above the resonance frequency. The level at lower frequencies (<4 kHz) decreases drastically as the source moves away from the hydrophone. The model and the measurements

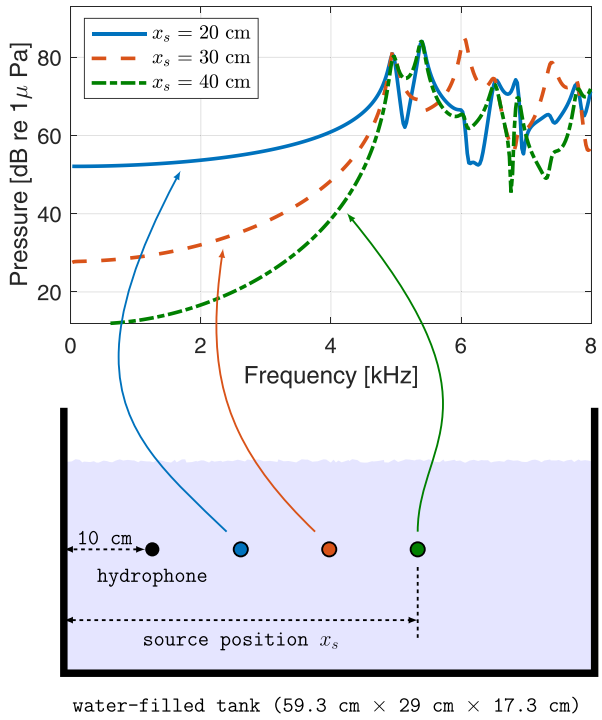


FIG. 2. (Color online) Theoretical frequency response functions (Green's functions) between the fixed hydrophone (black circle) and sources (colored circles). Three locations of the source are [20, 10, 10 cm] (blue), [30, 10, 10 cm] (red), and [40, 10, 10 cm] (green), respectively, and the location of the hydrophone is [10, 10, 10 cm].

provided in (Novak *et al.*, 2018) show a 30 dB level decrease per doubling of distance from the source. Consequently, the sound field created by a source, whose energy is distributed mainly below the first resonance frequency of the tank, will be strongly attenuated with the distance from the source. The reason for this behavior comes from the nature of the boundary conditions on the tank walls, which are close to the Dirichlet boundary conditions (reflection coefficient on the walls close to –1). This fact leads to emitted pressure waves that are reflected, in the low frequency range, with almost identical amplitude, but with an almost inverted phase, resulting in a quasi-cancellation of the direct and reflected waves.

The behavior at higher frequencies is different. Indeed, the level of the modes at and above the first resonance frequency (4.96 kHz, mode $m_x = 1$, $m_y = 1$, $m_z = 1$) varies greatly with the position of the source, and their level remains high compared to the ones below the resonance frequency. Audio samples, auralized using the model [Eq. (7)], are available online (Novak, 2019a) to illustrate the acoustic behavior of the fish tank. The algorithm of localization described in the following takes advantage of the detailed knowledge of the modal behavior of the fish tank.

III. SOUND RECORDING OF CROAKING GOURAMI (*TRICHOPSIS PUMILA*)

A. Measurement setup

The recordings of the fish sounds were performed in a rectangular water-filled tank of dimensions [$L_x = 59.3$ cm × $L_y = 29$ cm × $L_z = 17.3$ cm], L_z being the water height. The glass walls of the tank are 2.7 mm thick. Four Brüel & Kjær Hydrophones type 8103 are set inside the tank (see Table I for the coordinates). The signals from the hydrophones are amplified using a Nexus charge amplifier type 2692-A and acquired using Matlab and an RME Fireface 400 interface sampling at 192 kHz (see Fig. 3).

TABLE I. Tank dimensions and coordinates of the hydrophone positions used during the fish recordings.

	x (cm)	y (cm)	z (cm)
Tank dimensions	59.3	29.0	17.3
Hydrophone 1	8.2	11.0	4.0
Hydrophone 2	8.8	22.1	11.0
Hydrophone 3	43.5	22.0	12.2
Hydrophone 4	48.1	7.2	4.0

Six croaking gouramis (*Trichopsis pumila*) are placed in the water-filled tank. The Trichopsis species can produce impulsive sounds consisting of several bursts, each burst comprising one or two pulses (Daugherty and Marshall, 1976).

Any equipment likely to generate disturbing noises, such as filters or water pumps, is switched off during the recordings to increase the signal-to-noise ratio. The experiment takes place in a quiet room and the tank is placed on a 10 cm thick foam to eliminate the influence of external vibrations.

Two video recording systems are used during the measurements to validate the acoustic localization. The first video system is a Microsoft Kinect Sensor V2, placed 20 cm above the tank that uses an image processing technique with Snell's law correction to acquire the three dimensional (3D) position of each fish (Saberioon and Cisar, 2016). The Kinect system is used to verify the accuracy of the fish positions determined by the acoustic localization technique developed in this paper. The second video system is a conventional web camera, placed next to the Kinect sensor; it records a color video sequence that is further used to analyze fish behavior. Finally, all clocks of the measurement systems (acoustics, Kinect, and web camera) are synchronized.

B. Analysis of sound recordings

The audio recordings are analyzed to detect the sound events corresponding to the impulsive sounds created by

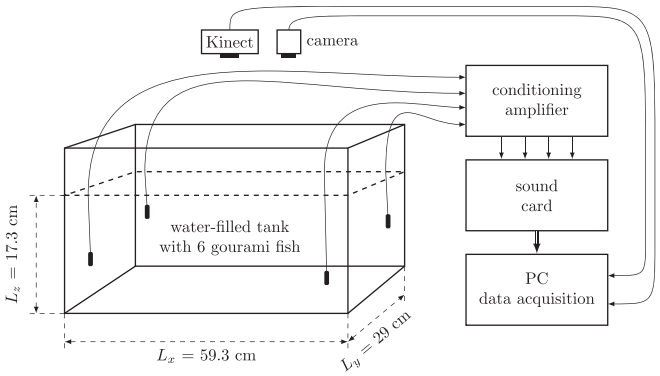


FIG. 3. Schematic representation of the measurement setup.

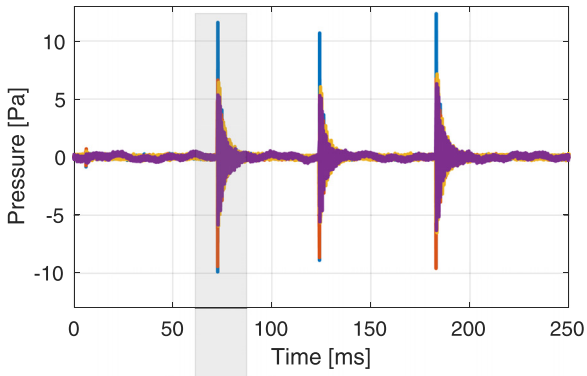
croaking gourami. In Fig. 4, one of the detected sequences of bursts and its spectrum are shown as an example. The waveform depicted in Fig. 4(a) consists of three bursts.

The recorded signal is a convolution product between the sound pressure generated by gourami and the impulse response of the fish tank (time-domain Green's function between the source and the receiver [Bruneau and Scelo (translator and contributor), 2006)]. Since the impulse response is known to be very long in time due to the reverberant acoustic space (Novak *et al.*, 2018), the recorded sounds contain long reverberant signatures as seen in Fig. 4(a).

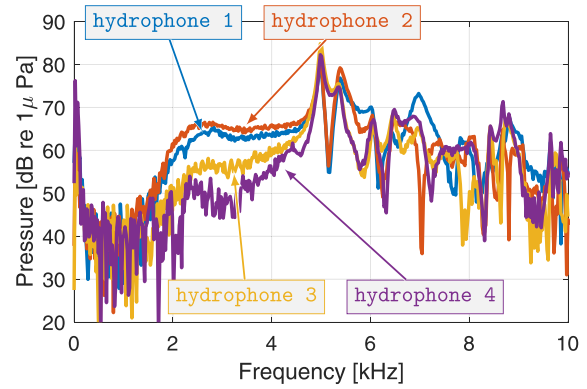
Figure 4(b) shows the spectra of one of the bursts measured by the four hydrophones. In the spectra, there are two very distinct frequency bands as predicted by the model. Below the first resonance frequency, the spectra of signals recorded by hydrophones contain energy (between 2 and 4.5 kHz) with level varying a lot from one hydrophone to the other. For instance, levels from hydrophones 3 and 4 are from 10 to 15 dB lower than those from hydrophones 1 and 2, indicating that the fish that produced the sound was closer to hydrophones 1 and 2. Above 4.5 kHz, the spectrum of all four hydrophones shows many resonances and anti-resonances (at least 10 up to 8 kHz) with almost the same amount of energy but with a varying shape of the spectra.

Finally, Fig. 4(c) shows a zoom of the beginning of the burst time waveforms received by the four hydrophones. It reveals that the signals from hydrophones 1 and 2 arrive approximately 0.1 ms before the signal from hydrophone 4. The underwater speed of sound being 1480 m/s, this time lag corresponds to a distance of 15 cm. In other words, the source emitting the sound must have been closer to the location of hydrophones 1 and 2 by approximately 15 cm than to hydrophone 4. This is correlated with the lower part of the spectra [up to 4.5 kHz, Fig. 4(b)] where the signals from hydrophones 1 and 2 contain more energy than the one from hydrophone 4.

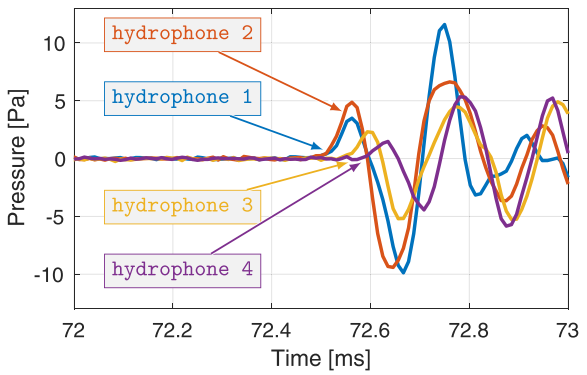
Note that for a sampling frequency 96 kHz, the time sampling period is about 0.01 ms, which corresponds to a distance of 1.5 cm. Since the beginning of the sound can only be estimated with time accuracy of a few samples, the methods based on TDOA can only lead to a rough estimation of the source position.



(a) example of the recorded croaking sound



(b) spectra of the selected burst



(c) zoom to the burst beginning

FIG. 4. (Color online) Example of recorded sounds of a croaking gourami: (a) waveform of four hydrophones; (b) corresponding spectra of a selected burst; (c) zoom on the beginning of the selected burst.

IV. LOCALIZATION ALGORITHM

The spectrum $P_n(\omega)$ of the sound pressure recorded with the hydrophone n can be expressed as

$$P_n(\omega) = G(\vec{r}_n, \vec{r}_s, \omega)X(\omega), \tag{8}$$

where $G(\vec{r}_n, \vec{r}_s, \omega)$ is the Green's function calculated between the position of the hydrophone n [$\vec{r}_n = (x_n, y_n, z_n)$], and the source position $\vec{r}_s = (x_s, y_s, z_s)$, and where $X(\omega)$ is the spectrum of the sound pressure emitted by the fish located at $\vec{r}_s$.

By dividing the spectrum $P_n(\omega)$ by the Green's function $G(\vec{r}_n, \vec{r}_s, \omega)$, one should, in a perfect case, access the spectrum $X(\omega)$ of the original sound pressure produced by the fish. However, due to the highly reverberant behavior of the fish tank, which has very narrow resonance peaks with high amplitudes and high quality factors, a small difference between the model and the actual fish tank behavior leads to a high inaccuracy in the deconvoluted spectrum $X(\omega)$.

The localization algorithm used in this paper is based on a ratio of two spectra ratios: the spectra ratio between the sound pressure measured by hydrophones at two locations and the spectra ratio between the theoretical Green's functions at the same locations. The comparison is provided in the frequency band corresponding to the first few (e.g., ten) modes that are between 4 and 8 kHz as shown in Fig. 2 (note that the frequency band is closely related to the size of the tank). To give an example of the spectra ratio shape in this frequency band, we compare in Fig. 5 the spectra ratio

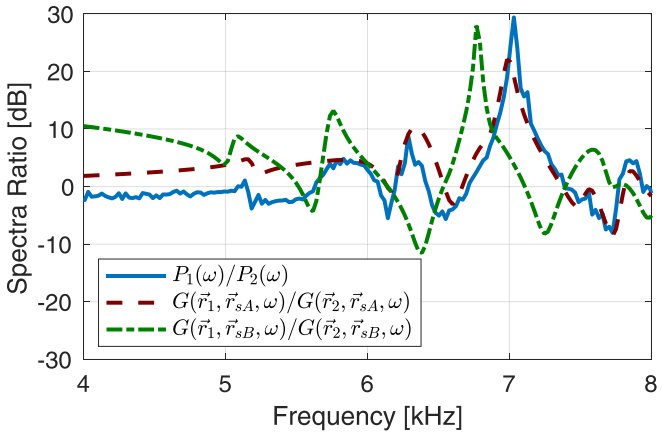


FIG. 5. (Color online) Comparison of the spectra ratio $P_1(\omega)/P_2(\omega)$ of the selected burst recorded by hydrophones 1 and 2 (in blue solid line) with the ratio of estimated Green's function, the source being placed in positions [22.5, 18.0, 6 cm] (red dashed line) and [5, 10, 2 cm] (green dashed-dotted line).

$P_1(\omega)/P_2(\omega)$ of the recorded sound pressures (solid blue line) with two Green's functions ratio $G(\vec{r}_1,\vec{r}_{sA},\omega)/G(\vec{r}_2,\vec{r}_{sA},\omega)$ and $G(\vec{r}_1,\vec{r}_{sB},\omega)/G(\vec{r}_2,\vec{r}_{sB},\omega)$, the first one chosen at the source position $\vec{r}_{sA}=[22.5, 18, 6\,\text{cm}]$ (dashed red line) and the second one at the source position $\vec{r}_{sB}=[5, 10, 2\,\text{cm}]$ (dashed-dotted green line). The spectra ratio $P_1(\omega)/P_2(\omega)$ is then much more similar to the Green's functions ratio at the source position $\vec{r}_{sA}$ compared to the one at $\vec{r}_{sB}$, indicating that the position of the fish is closer to the position $\vec{r}_{sA}$.

In the following, the spectra ratio $S_{m,n}(\vec{r},\omega)$ is defined as a criterion for the localization algorithm:

$$S_{m,n}(\vec{r},\omega)=\frac{G(\vec{r}_m,\vec{r},\omega)}{G(\vec{r}_n,\vec{r},\omega)}\frac{P_n(\omega)}{P_m(\omega)}. \tag{9}$$

In a perfect case, the source being perfectly localized and the Green's functions perfectly describing the real fish tank, the spectra ratio $S_{m,n}(\vec{r},\omega)$ is equal to 1 [inserting Eq. (8) to Eq. (9)]. In a real-world application the closer the ratio $S_{m,n}(\vec{r},\omega)$ is to one, the higher the probability that the candidate source position is the real source position. To estimate the closest match, we calculate the root mean square of $S_{m,n}(\vec{r},\omega)$ in dB scale as

$$\sigma_{m,n}(\vec{r})=\sqrt{\frac{1}{\omega_2-\omega_1}\int_{\omega_1}^{\omega_2}\left(20\log_{10}|S_{m,n}(\vec{r},\omega)|\right)^2 d\omega}, \tag{10}$$

where ω_1 and ω_2 are chosen to cover the frequency band from 4 to 8 kHz. Last, the detection factor $DF(\vec{r})$ is a function that takes into account all the functions $\sigma_{m,n}(\vec{r})$ of four hydrophones as

$$DF(\vec{r})=\frac{1}{\prod_{m=1}^{3}\prod_{n=m+1}^{4}\sigma_{m,n}(\vec{r})}. \tag{11}$$

A. Dictionary of Green's functions

The searching for the most probable source location $\hat{\vec{r}}_s=(\hat{x}_s,\hat{y}_s,\hat{z}_s)$ that maximizes $DF(\vec{r})$ can be done either by the computation of a dictionary of Green's functions in a given grid or by using a searching method such as genetic algorithm (Goldberg, 1989) that calculates the Green's functions in each step.

The generation of a dictionary of Green's functions on a given grid takes more time compared to a searching method (the time consumption of both methods is discussed in Sec. V). Nevertheless, calculating the dictionary of Green's function allows to plot a detection factor map which provides a visual information about all the tested candidate source positions in a two dimensional (2D) colored plot. For this reason, we present the results of the localization algorithm based on the dictionary method and we provide an overview of an optimization using genetic algorithm in Sec. V).

The first step of the dictionary-based algorithm consists of computing the dictionary of Green's functions in a x, y, z grid with a given step (0.5 cm in our example). For each grid, four sets of Green's function $G(\vec{r}_n,\vec{r},\omega)$ $(n=1,\ldots,4)$ between the hydrophone n position $\vec{r}_n=(x_n,y_n,z_n)$ and each point $\vec{r}=(x,y,z)$ in the grid are computed using Eq. (7). This step is done once for the given setup of the tank dimensions and hydrophone positions (see Table I).

The spectra $P_n(\omega)$, $n\in(1,\ldots,4)$ of a selected burst of recorded sound pressure emitted by the fish is then used to compute the detection factor $DF(\vec{r})$ for each tested candidate source position in the grid using Eqs. (9)–(11).

B. Results

The algorithm described above is applied to three selected recordings of the sound events produced by croaking gourami (see Sec. III) to localize their position at the time they produced sounds. To verify that each estimated position is related to a real fish position in the tank, we compare the acoustic-based localization with the one from the Kinect system (Saberioon and Cisar, 2016) and also visually with the video recordings.

1. Acoustic localization

The most probable source location $[\hat{x}_s,\hat{y}_s,\hat{z}_s]$ estimated from Eq. (11) is provided in Table II. A 2D map $DF(\vec{r})$ for $\vec{r}=(x,y,z=\hat{z}_s)$ is depicted for each of the three tested fish sounds in Figs. 6, 7, and 8. The dark color represents the highest value of the detection factor $DF(\vec{r})$, the bright color represents the lowest one.

TABLE II. Results of the acoustic-based localization algorithm compared with the localization using the image processing technique for three tested sound events.

	Estimated position		
Time (figure)	By acoustics $\hat{x}_s,\hat{y}_s,\hat{z}_s$ (cm)	From Kinect x_s,y_s,z_s (cm)	$\lvert\hat{\vec{r}}_s-\vec{r}_s\rvert$ (cm)
16:16:12 (Fig. 6)	35.0, 5.5, 6.5	34.9, 6.6, 2.5	0.1, 1.0, 4.0
16:36:26 (Fig. 7)	52.5, 18.5, 9.0	53.9, 19.5, 2.8	1.4, 1.0, 6.2
19:59:09 (Fig. 8)	22.5, 18.0, 6.0	20.5, 18.6, 2.7	2.0, 0.6, 3.3

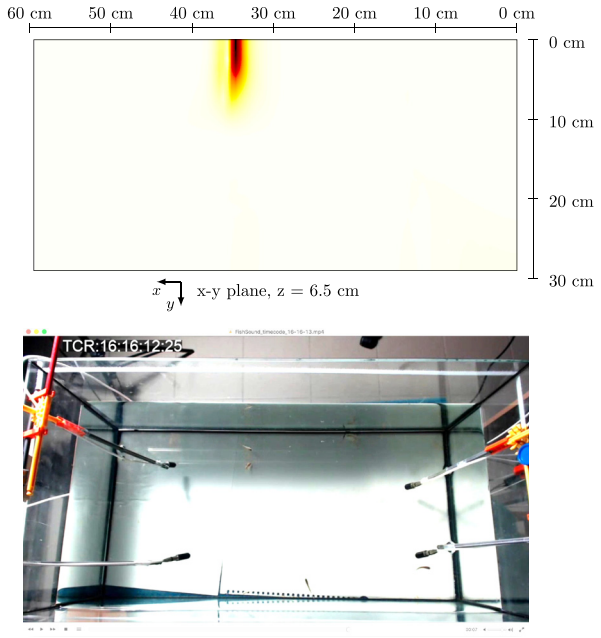


FIG. 6. (Color online) A $(x-y)$ surface plot ($z = 6.5$ cm) of the DF map for the sound event appearing in time frame 16:16:12.25 (above), and the print-screen of the video recording at the same time instant.

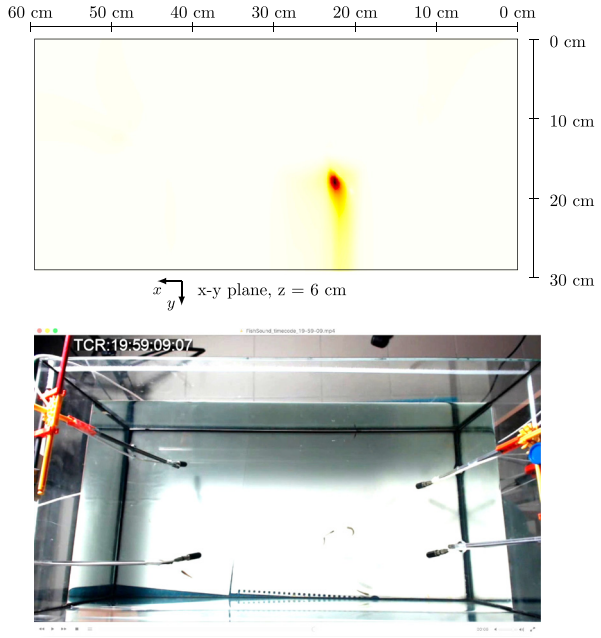


FIG. 8. (Color online) A $(x-y)$ surface plot $x-y$ ($z = 6$ cm) of the DF map for the sound event appearing in time frame 19:59:09.07 (above), and the print-screen of the video recording at the same time instant.

2. Kinect localization

The system based on image processing from the data from Kinect sensor provides the coordinates of all the fish in the tank (Saberioon and Cisar, 2016). Position of the fish closest to the position $\vec{r}_s = (\hat{x}_s, \hat{y}_s, \hat{z}_s)$ at the time of the sound event is provided in Table II.

3. Video recording

Recordings [provide online at Novak (2019b)] can be used to visually verify that the position detected by both image and acoustic methods is correct and can also bring information about the behavior of the fish at the time of the sound event.

A print-screen of the video recording at the time corresponding to the sound event is provided below the detection factor maps in Figs. 6, 7, and 8. A brief behavior analysis is discussed in Sec. V.

For each tested sound event, the acoustic localization algorithm provides an estimated position that is very close to one of the fish-positions localized by the Kinect system. Since the Kinect system can only determine the position of each fish in the tank but cannot predict which fish is sound producer, we use the video recordings that can provide more visual information. It shows that during each sound event, there is a fish couple at the location found by the algorithm that changes the behavior at the moment of the sound event. The video recordings are provided online (Novak, 2019b) (print-screens provided in Figs. 6, 7, and 8).

The first tested sound event at time of recording 16:16:12 (Fig. 6) shows the maximum in the detection factor map $DF(\vec{r})$ at the coordinates [35.0, 5.5, 6.5]. The closest fish detected by the Kinect system is at the coordinates [34.9, 6.6, 2.5]. The acoustic localization algorithm estimates the position of a source emitting the second tested sound event at a time of recording 16:36:26 (Fig. 7) at [52.5, 18.5, 9.0] and the closest fish detected by the Kinect system is at the coordinates [53.9, 19.5, 2.8]. Finally, the position of source emitting the third tested sound event at time of recording 19:59:09 (Fig. 8) is estimated at [22.5, 18.0, 6.0], the closest fish being detected by the Kinect system at the coordinates [20.5, 18.6, 2.7].

The difference between the acoustic and Kinect system localization is lower than 2.1 cm in $x-y$ plane for all tested sound events. The estimation in $z-$coordinate is less precise; this is discussed in Sec. V.

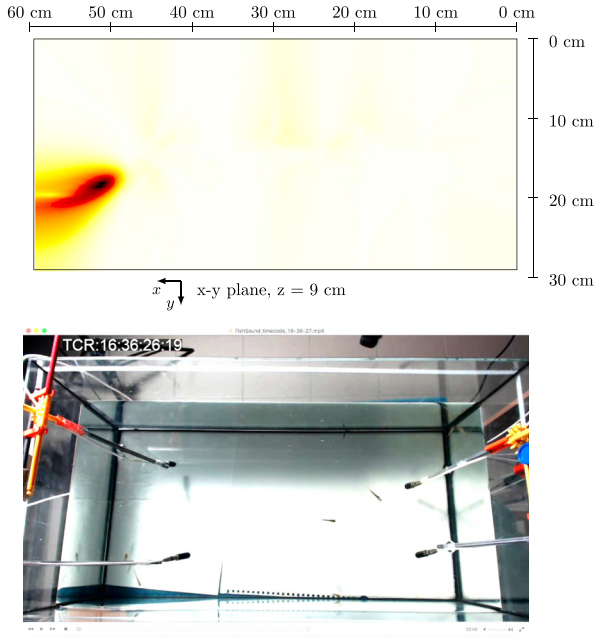


FIG. 7. (Color online) A $(x-y)$ surface plot ($z = 9$ cm) of the DF map for the sound event appearing in time frame 16:36:26.19 (above), and the print-screen of the video recording at the same time instant.

C. Localization with two hydrophones

It has been shown that using four hydrophones leads to satisfying results. In practice, it is desired to use as few hydrophones as possible to minimize the cost and complexity of the implementation. For a pair of hydrophones, the detection factor $DF_{m,n}(\vec{r})$ can be estimated as

$$DF_{m,n}(\vec{r}) = \frac{1}{\sigma_{m,n}(\vec{r})}. \tag{12}$$

In Fig. 9, the functions $DF_{m,n}(\vec{r})$ for $\vec{r} = (x, y, z = \hat{z}_s)$ are depicted for each pair of hydrophones separately for the last of the previously tested cases (localized at the coordinates [22.5, 18.0, 6.0]). While the results are less precise than in the case of localization with four hydrophones, the detection factor maps $DF(\vec{r})$ in Fig. 9 show that the localization with only two hydrophones leads to an approximate position of the fish depending on the pair of the chosen hydrophones. For example, results from the hydrophones 1 and 2 [Fig. 9(a)] gives a local maximum at the same position that the one obtained with a localization based on four hydrophones while the localization by the hydrophones 3 and 4 [Fig. 9(f)] provides a less focused result with maximum shifted towards the fish tank wall by few centimeters.

V. DISCUSSION

A. Time consumption of the algorithm

The results shown in Sec. IV are obtained using the dictionary of Green's functions that is computed in a 3D grid with a step of 0.5 cm, which makes 232 thousand candidate source positions. As four hydrophones were used and the set of Green's functions is computed for each hydrophone, the dictionary contains almost a million Green's functions. Each Green's function is computed at 150 frequency bins in the given frequency region from 4 to 8 kHz within 2 ms with an ordinary personal laptop. The preparation of the dictionary takes approximately 30 min, and the comparison of the recorded spectra with each point in the dictionary takes approximately 20 min.

However, these high computation times can be largely reduced using a searching method such as genetic algorithm. We tested a genetic algorithm (Goldberg, 1989), for which the number of offspring (number of tested Green's functions

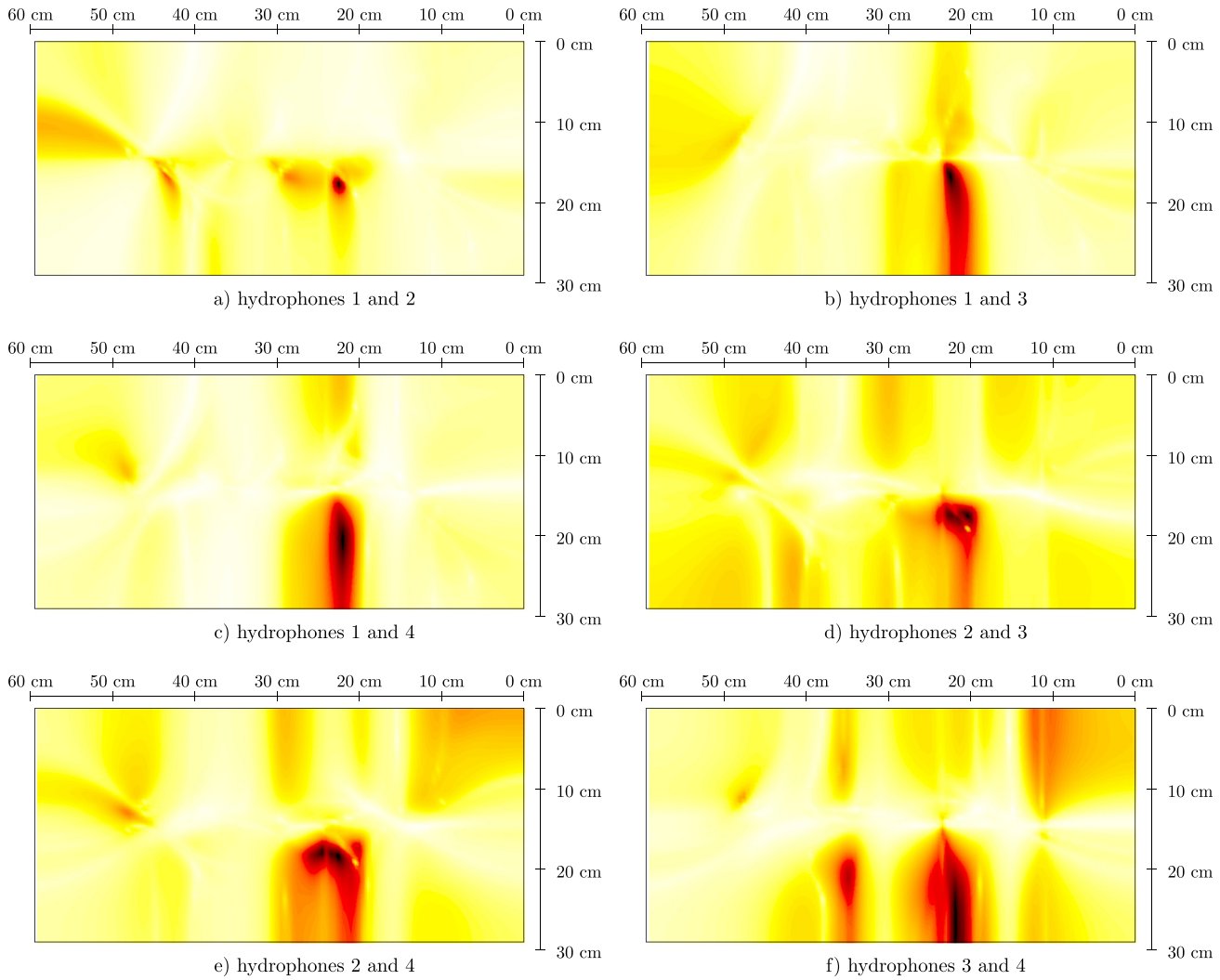


FIG. 9. (Color online) A (x – y) surface plot (z = 6 cm) of the DF maps of the fish position using two hydrophones: (a) hydrophones 1 and 2, (b) hydrophones 1 and 3, (c) hydrophones 1 and 4, (d) hydrophones 2 and 3, (e) hydrophones 2 and 4, (f) hydrophones 3 and 4.

in one generation of the algorithm) is set to 100 and number of parents (number of selected Green's functions for next generation) is set to ten. Within five generations the algorithm converged to the position that was found by the dictionary algorithm for all tested cases. Since only 500 Green's functions are generated and tested with the genetic algorithm, the computation time reduces to 3 s.

B. Precision of localization

The acoustic-based localization algorithm presented above shows that each of the coordinates associated with a sound event was very close to at least one fish detected by the Kinect system. Considering the image processing system as a reference system, the localization error of the acoustic localization in the $x-y$ plane is within 2 cm. The precision of localization in the $z-$ direction is much lower (see Table II), which may be due to the difference between the model and the real setup. Indeed, the acoustic fish tank model considers the five glass-walls to be surrounded by air, including the bottom wall. In the measurement conditions, the tank was supported by a 10 cm thick foam whose acoustic properties might be different from those of air. This inaccuracy of the boundary conditions at the bottom wall can lead to lower precision of the localization algorithm in the z-direction.

C. Constrains for a successful localization

Several hypotheses were put forward in the paper. First, we consider that the analytical fish tank model (Novak *et al.*, 2018) describes perfectly the fish tank. As shown in this paper the model is very accurate, and the resonance frequencies are matched within a 2% error. However, the presence of the fish in the tank may influence the precision of the model. As shown in the present paper, the presence of six croaking gouramis (each of them being 1–2 cm long) did not disturb the localization procedure. Tests with bigger sound-producing fish and with increased number of fish are planned to test the influence of the fish body on the model.

Next, the model considers that an omnidirectional point source produces the sound, while, in the real-world application, the fish body may influence the directivity of the source. The small level difference in Fig. 5 between the measured spectra ratio (solid blue line) and the model-based spectra ratio (dashed dark red line) may come from the possible non-uniform directivity pattern of the sound-producing fish. Nevertheless, on the one hand, this does not seem to influence the localization algorithm, and on the other hand, there is no study, to our knowledge, dealing with the directivity of the croaking gourami that could be included into the algorithm.

Finally, the localization is based on spectra comparison in the frequency band of the first few resonance frequencies of the tank that, in the case of a tank with dimensions $59.3\,\text{cm} \times 29\,\text{cm} \times 17.3\,\text{cm}$, are located between 4 and 8 kHz. It is thus necessary that the fish can excite this frequency region [see Fig. 4(b)].

VI. CONCLUSION

In this paper, we develop an algorithm to localize sound-producing fish placed in a small rectangular tank. Such a technique may be useful for bioacoustic behavioral studies to determine which fish in a group are sound producers and, if video recording is provided, in which behavioral context the sound was produced.

The proposed method is tested on real data gathered using four hydrophones placed in a rectangular fish tank together with a group of sound-producing fish (six croaking gouramis). The algorithm uses a recently developed analytical model of a fish tank and is based on the comparison of the ratio of the recorded and model-based spectra in the frequency region of the first few resonance frequencies of the tank. A dictionary of Green's functions computed by the analytical model is used in the paper to estimate the probability of the position of the sound source computed in a 3D grid to provide visual information of the localization. Optimization based on a genetic algorithm that reduces the computation time is also discussed.

The results from the acoustic localization are compared with video recordings and with a localization technique based on image processing. The precision of the acoustic localization, compared with the image processing based localization, is within 2 cm in the $x-y$ plane.

ACKNOWLEDGMENTS

This research was funded by the Region Pays de la Loire within the Le Mans Acoustic Project and received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No. 652831 (AQUAEXCEL^{2020}). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein. The authors are indebted to Christoph Steinbach (Faculty of Fisheries and Protection of Waters, JCU, Czech Republic) for behavioral analysis and Marcel Koken (LABOCEA, Brest, France) for substantial help in providing, respectively, the early subject matter of the project, and Eric Bavu (CNAM, Paris, France) for helpful discussions. They express their gratitude to Louis Lotton, the author of Fig. 1.

APPENDIX

This appendix provides a robustness analysis of the position of hydrophones that may be useful for the practical application of the method described in this document. Since the Green's function of a rectangular water-filled tank contains many low-damping resonances, it could be suspected that the effect of slight changes in the hydrophone position may influence robustness.

We run a simulation in which the tank dimensions and hydrophone positions were taken from Table I and a Dirac-like source was placed in an arbitrary position [15, 16, 6 cm]. The spectra $P_n(\omega)$ were simulated using Eq. (8) in which $X(\omega) = 1$ (Dirac-like source) and the Green's functions $G(\vec{r}_n, \vec{r}_s, \omega)$ were calculated using [Eq. (7)]. Hydrophone positions $\vec{r}_n$ were randomly selected from a normal distribution with mean value equal to the real position of each hydrophone

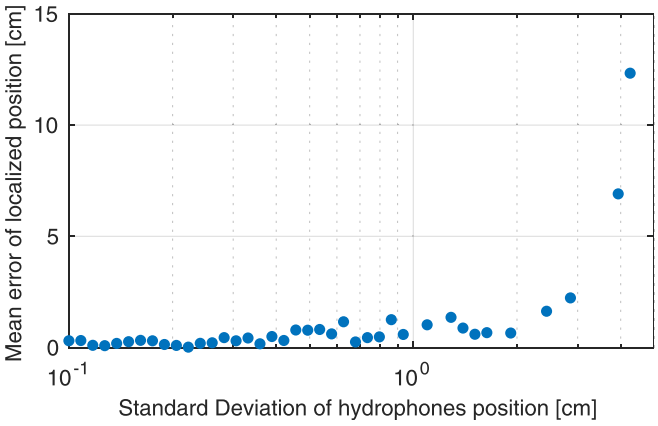


FIG. 10. (Color online) Mean error of the estimated position as a function of the error of the hydrophone positions.

and standard deviation σ_r. We tested 40 different standard deviations σ_r going from 1 mm to 5 cm, each standard deviation being tested for 200 random hydrophone positions. We used the genetic algorithm described above for the localization.

The mean absolute error of the localized position is depicted in Fig. 10. One can conclude that if the real hydrophone position is shifted from its denoted position with a standard deviation not exceeding 3 cm, the mean error of the localized position remains less than 2 cm. Exceeding the 3 cm standard deviation in the hydrophone position leads to a non-acceptable error.

Akamatsu, T., Okumura, T., Novarini, N., and Yan, H. Y. (**2002**). "Empirical refinements applicable to the recording of fish sounds in small tanks," J. Acoust. Soc. Am. **112**(6), 3073–3082.

Asano, F., Goto, M., Itou, K., and Asoh, H. (**2001**). "Real-time sound source localization and separation system and its application to automatic speech recognition," in *Seventh European Conference on Speech Communication and Technology*, Aalborg, Denmark.

Bertucci, F., Beauchaud, M., Attia, J., and Mathevon, N. (**2010**). "Sounds modulate males aggressiveness in a cichlid fish," Ethology **116**(12), 1179–1188.

Bruneau, M., and Scelo (translator and contributor), T. (**2006**). *Fundamentals of Acoustics* (ISTE, London).

Chardon, G., Nowakowski, T., De Rosny, J., and Daudet, L. (**2015**). "A blind dereverberation method for narrowband source localization," IEEE J. Sel. Top. Signal Process. **9**(5), 815–824.

Daku, B. L. F., Salt, J. E., and McIntyre, C. M. (**1992**). "Quality of underwater source localization in a multipath environment," J. Acoust. Soc. Am. **91**(2), 957–964.

Daugherty, J. L., and Marshall, J. A. (**1976**). "The sound-producing mechanism of the croaking gourami, *trichopsis vittatus* (pisces, belontiidae)," Physiol. Zool. **49**(2), 227–244.

Dokmanić, I., and Vetterli, M. (**2012**). "Room helps: Acoustic localization with finite elements," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2617–2620.

Fay, R. R. (**2008**). "Fish bioacoustics," in *Handbook of Signal Processing in Acoustics* (Springer, New York), pp. 1851–1860.

Fink, M., and de Rosny, J. (**2001**). "Time-reversed acoustics in random media and in chaotic cavities," Nonlinearity **15**(1), R1–R18.

Fink, M., Montaldo, G., and Tanter, M. (**2004**). "Time reversal acoustics," in *2004 IEEE Ultrasonics Symposium* (IEEE, New York), Vol. 2, pp. 850–859.

Goldberg, D. E. (**1989**). *Genetic Algorithms in Search, Optimization, and Machine Learning* (Addison-Wesley Longman Publishing Co., Inc., Boston, MA).

Hawkins, A. (**1986**). "Underwater sound and fish behaviour," in *The Behaviour of Teleost Fishes* (Springer, New York), pp. 114–151.

Johnson, D. H., and Dudgeon, D. E. (**1993**). *Array Signal Processing: Concepts and Techniques* (PTR Prentice Hall, Englewood Cliffs, NJ).

Kim, S.-M., Oh, S., and Byun, S.-H. (**2015**). "Underwater source localization in a tank with two parallel moving hydrophone arrays," in *OCEANS'15 MTS/IEEE Washington*, (IEEE, New York), pp. 1–4.

Ladich, F. (**2015**). *Sound Communication in Fishes* (Springer, New York), Vol. 4.

Malioutov, D., Cetin, M., and Willsky, A. S. (**2005**). "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Trans. Signal Process. **53**(8), 3010–3022.

Novak, A. (**2019a**). "Acoustics of a fish-tank," ant-novak.com/posts/research/2018-07-29_AAA_FishTank/ (Last viewed 2019-02-10).

Novak, A. (**2019b**). "Localization of moving acoustic sources in a water-filled tank," ant-novak.com/posts/research/2019-02-10_JASA_FishLocalisation/ (Last viewed 2019-02-10).

Novak, A., Bruneau, M., and Lotton, P. (**2018**). "Small-sized rectangular liquid-filled acoustical tank excitation: A modal approach including leakage through the walls," Acta Acust. Acust. **104**(4), 586–596.

Nowakowski, T., de Rosny, J., and Daudet, L. (**2017**). "Robust source localization from wavefield separation including prior information," J. Acoust. Soc. Am. **141**(4), 2375–2386.

Saberioon, M., and Cisar, P. (**2016**). "Automated multiple fish tracking in three-dimension using a structured light sensor," Comput. Electron. Agr. **121**, 215–221.

Stefanoiu, D., and Ionescu, F. (**2003**). "A genetic matching pursuit algorithm," in *2003 IEEE Seventh International Symposium on Signal Processing and Its Applications* (IEEE, New York), Vol. 1, pp. 577–580.

Zelick, R., Mann, D. A., and Popper, A. N. (**1999**). *Acoustic Communication in Fishes and Frogs* (Springer, New York), pp. 363–411.