

Time of arrival difference estimation for narrow band high frequency echolocation clicks

Douglas Gillespie, and Jamie Macaulay

Citation: *The Journal of the Acoustical Society of America* **146**, EL387 (2019); doi: 10.1121/1.5129678

View online: <https://doi.org/10.1121/1.5129678>

View Table of Contents: <https://asa.scitation.org/toc/jas/146/4>

Published by the [Acoustical Society of America](https://www.asa.org/)

ARTICLES YOU MAY BE INTERESTED IN

[Three-dimensional acoustic propagation effects induced by the sea ice canopy](#)

The Journal of the Acoustical Society of America **146**, EL364 (2019); <https://doi.org/10.1121/1.5129554>

[Deep transfer learning for source ranging: Deep-sea experiment results](#)

The Journal of the Acoustical Society of America **146**, EL317 (2019); <https://doi.org/10.1121/1.5126923>

[Fidelity of low-frequency underwater acoustic measurements by sensors mounted on compact platforms](#)

The Journal of the Acoustical Society of America **146**, EL405 (2019); <https://doi.org/10.1121/1.5130755>

[Acoustic end corrections for micro-perforated plates](#)

The Journal of the Acoustical Society of America **146**, EL399 (2019); <https://doi.org/10.1121/1.5129560>

[Broadband sound propagation in a seagrass meadow throughout a diurnal cycle](#)

The Journal of the Acoustical Society of America **146**, EL335 (2019); <https://doi.org/10.1121/1.5127737>

[Effects of temporal distortions on consonant perception with and without undistorted visual speech cues](#)

The Journal of the Acoustical Society of America **146**, EL381 (2019); <https://doi.org/10.1121/1.5129562>



JASA
THE JOURNAL OF THE
ACOUSTICAL SOCIETY OF AMERICA

Special Issue:
Additive Manufacturing and Acoustics

Submit Today!

The banner features a background image of a 3D printer nozzle printing a red, lattice-structured part. The text is overlaid on the left side of the image.

Time of arrival difference estimation for narrow band high frequency echolocation clicks

Douglas Gillespie and Jamie Macaulay

*Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews,
 St Andrews, KY16 8LB, Scotland
 dg50@st-andrews.ac.uk, jdjm@st-andrews.ac.uk*

Abstract: Algorithms are presented for the accurate time of arrival difference estimation of high frequency narrow band echolocation clicks from Harbor Porpoise. These clicks typically have a center frequency of around 130 kHz (wavelength ~ 1.2 cm) and duration of < 0.1 ms. When using hydrophones spaced centimeters apart, spatial aliasing can cause large errors on inter-hydrophone timing measurements due to the incorrect peak in the cross-correlation function of two signals being selected. It is shown that at sample rates of less than about 6 times the fundamental frequency, the incorrect correlation peak will be selected in 55% of measurements leading to large errors in time of arrival estimates. For clicks with a SNR > 10 dB these errors can be reduced by over two orders of magnitude through a combination of up-sampling the data and parabolic interpolation of peaks in the cross-correlation functions.

© 2019 Acoustical Society of America

[LNK]

Date Received: July 10, 2019 **Date Accepted:** September 30, 2019

1. Introduction

Several species of cetacean, including the harbor porpoise (*Phocoena phocoena*), *Kogia*, and *Cephalorhynchus* dolphins all produce narrow band high frequency (NBHF) echolocation clicks with center frequencies at around 130 kHz and durations of around $100 \mu\text{s}$ (e.g., [Teilmann et al., 2002](#); [Kyhn et al., 2010](#)). Many applications of passive acoustic monitoring (PAM) for these species employ arrays of two or more hydrophones in order to localize sounds in 1, 2, or 3 dimensions by measuring time of arrival differences (TOADs) between the different hydrophones (e.g., [Macaulay et al., 2017](#)). A number of autonomous or vessel based recording systems are available that can sample at a high enough rate to capture these sounds with suitable systems typically having sample rates of between 312 and 576 kHz ([Sousa-Lima et al., 2013](#)). Particularly for autonomous systems, whose deployment time is likely to be limited by data storage capacity, there is an obvious incentive to sample at the lowest rate possible.

The narrow band nature of porpoise clicks can make the cross-correlation of a pair of signals prone to ambiguities due to spatial aliasing of the signal. Several peaks in the cross-correlation function of narrow band clicks have similar amplitudes, so noise and the effect of sampling the data at discrete intervals can lead to incorrect peak selection in the correlation function.

The fundamental limits of TOAD estimation for narrow band, ambiguity prone, signals are discussed in [Weiss and Weinstein \(1983\)](#) who present three zones of accuracy. In the first, SNR is so poor that no information can be obtained from the signals, uncertainty being defined by the range of possible values defined by the sensor separation. In the second zone, levels of noise make it impossible to determine which peak in the cross correlation function is correct so accuracy is determined by the duration of the signals envelope. In the third zone, at high SNR, the correct peak in the cross correlation function can be selected and accuracy becomes high.

With digitally sampled signals, the issues of cross correlation peak selection described by Weiss and Weinstein are exacerbated at low sample rates since, as the sample rate is reduced towards the Nyquist rate (i.e., $2\times$ the highest frequency component of the signal) the chances of sampling the correlation function close to its true peak become small compared to the chance of sampling close to the top of one of the many other peaks in that function. As an example, [Fig. 1\(A\)](#) shows a simulated porpoise click, with the waveform on channel B delayed by $6.8 \mu\text{s}$ (about 1 cm travel distance) compared to channel A. [Figure 1\(B\)](#) shows the true correlation function of the two waveforms and the actual points in the cross correlation function obtained when the waveforms are sampled at a rate of 500 kHz. By chance the highest value in the

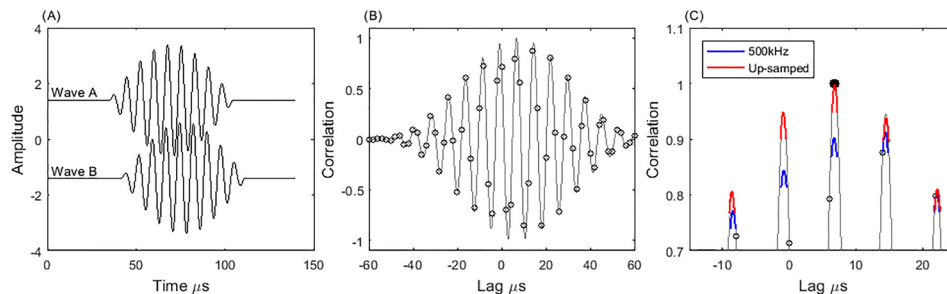


Fig. 1. (Color online) (A) Simulated porpoise waveform and (B) cross correlation function (solid line) and points on the correlation function for 500 kHz sample data (open circles). The filled circle is the correct correlation maximum. (C) Parabolic peak interpolation for 500 kHz data and the same 500 kHz data up-sampled to 1 MHz.

correlation function from the sampled data does not lie on the true peak of the cross correlation function, but on the next peak along, at a value of $14 \mu\text{s}$ (7 samples), an error of $7.2 \mu\text{s}$. Problems of peak selection become exacerbated in the presence of noise which causes random fluctuations in the height of each peak additional to the height measurement errors caused by sampling.

We define two types of error: “small errors” are measurement errors about the true peak in the cross correlation function, i.e., $<1/2$ cycle, so for porpoise clicks are $<3.8 \mu\text{s}$, “gross errors” occur when the TOAD measurement algorithm fails to find the correct peak in the cross correlation function, so are larger than $3.8 \mu\text{s}$. A number of relatively simple enhancements to the cross correlation algorithm can reduce the numbers of gross errors and significantly reduce the overall error on TOAD measurements. The three enhancements examined are parabolic peak interpolation of the cross correlation function, waveform envelope tracing and up-sampling of the waveform data.

The algorithms are intended for use in real time detection and localization systems, so execution speed is an important consideration. Algorithm choices have therefore been selected to relatively straightforward cross correlation methods that have been implemented in the PAMGuard software (Gillespie *et al.*, 2008). This is being used for continuous long term monitoring of a twelve hydrophone shore cabled array sampling each hydrophone at 500 kHz. During the first 12 months of data collection, over 740×10^6 transient sounds were detected and localized.

2. Methods

Porpoise like signals were simulated as a pure sine wave centered at 130 kHz windowed with a sine function shaped window of length $77 \mu\text{s}$ (Teilmann *et al.*, 2002). A sine window function rather than the often used raised cosine or Hann window was used since this qualitatively better reflected the sharper onset and offset of porpoise clicks than the raised cosine function and gave a -3 dB bandwidth for the generated clicks of 15.3 kHz, which is closer to the 16 kHz -3 dB bandwidth reported by Teilmann *et al.* (2002) than the 18 kHz bandwidth obtained when using a raised cosine window.

Each pair of clicks was generated with a random time of arrival difference (TOAD) between $\pm 100 \mu\text{s}$, equivalent to 15 cm travel distance in water. The start times of both clicks of each pair were offset by a random fraction of a sample to ensure that clicks were not always sampled at the same points in time. All clicks were generated with a zero to peak amplitude of 2 units, which gives an rms level within the envelope of each click of exactly 1. Varying amounts of incoherent white noise were then added to the signals. For each modification of the algorithm, 1000 porpoise signals were simulated at a sample rate of 500 kHz.

2.1 Time delay measurements

A number of enhancements to a basic cross correlation of the two waveforms were investigated in order to determine which produced the most accurate measurement of the correct time delay.

Filtering. Filtering to remove noise from parts of the spectrum not used by harbor porpoise is an obvious method of improving signal to noise ratio. Filtering can be conducted in the frequency domain during the cross correlation process by restricting the calculations to the frequency band of interest and therefore has the potential to not only improve TOAD measurements, but also to reduce processing load. All methods tested filtered the data using a 6 pole Butterworth band pass 100–150 kHz filter.

Parabolic peak interpolation. A common method for determining both the position and the height of a peak in many types of sampled data is parabolic interpolation [e.g., image processing—Fisher and Naidu (1996) and spectral peak determination—Gasior (2004)]. The method is simple: From the Taylor expansion of the cosine function, close to its peak $\cos(x) \approx 1 - x^2/2$. The highest point close to a peak is taken along with one point to either side. A parabola is fitted to the three points and both the position and the height of the highest point calculated. By better estimating the height of each peak between samples in the correlation function it is more likely that the correct peak will be selected. The improvement in timing accuracy, once the correct peak has been selected, is also directly beneficial. In our method, we calculate the height of the parabolic peak for all peaks within the range of possible time delays in a cross correlation function and select the highest.

Up-sampling. A parabolic fit about a sinusoidal peak will become poor when any of the points used in that fit are taken from the lower half of the sinusoidal function, i.e., beyond the point at which the curvature of the signal changes sign as it crosses zero. Ensuring that at least three positive valued samples can be taken from a function requires that the sample rate is at least six times the fundamental frequency of the data. While sampling at higher data rates may often be impractical for reasons of data storage, it is possible to up-sample data during the measurement of time delays by interleaving the waveform with zeros and then filtering with a low pass filter with a cut off frequency at or below the Nyquist frequency of the original data. Up-sampling has the advantage over parabolic interpolation in that it is fitting a sinusoidal model, rather than a parabolic model to the data, thereby maintaining its accuracy further from the correlation peak. In our method, we use both: first up-sampling, then fitting parabolic peaks. This is evident in Fig. 1(C) where parabolic peaks fitted to up-sampled data more accurately fit the true correlation function, whereas with 500 kHz sampled data, the parabolic fit around the incorrect peak at $14.5 \mu\text{s}$ remains higher which would lead to a gross error in the time delay measurement.

Envelope tracing. An alternative method which can be used to avoid the problem of selecting the correct correlation peak is to work with the envelope of the waveform, rather than the high frequency waveform itself. The envelope has a much lower fundamental frequency than the signal (around 6.5 kHz for a $77 \mu\text{s}$ duration signal). By using the signal envelope, it can be expected that although the chances of selecting the incorrect peak in the correlation function are reduced, the uncertainty in the position of that peak is likely to increase. In our method, the signal envelope is used in a cross correlation function along with parabolic interpolation.

2.2 Porpoise pinger trial data

As well as simulated data, the algorithms were tested on data collected with a four-channel recording device (SoundTrap Ocean Instruments, Ltd.). Hydrophones were in a tetrahedral configuration with 5 cm separation and a sample rate of 384 kHz per channel. The device was deployed for 4 days on a gill net south of Falmouth Bay, Cornwall, UK.

2.3 Theoretical bounds

If interpolation of the signal is not used, accuracy is limited to ± 0.5 samples. This is equivalent to an rms error of $1/\sqrt{12}$ samples (from $\int_{-0.5}^{+0.5} x^2 dx$), which for 500 kHz sampled data is $\pm 0.58 \mu\text{s}$.

Where accuracy is not limited by sampling resolution, the Cramer-Rao lower bound on timing accuracy is discussed for narrow band pulses by Weiss and Weinstein (1983). If the correct peak in the correlation function can be found, the Cramer-Rao lower bound can be approximated to

$$\text{var}(\delta t) = \frac{1}{2\Delta f T \text{SNR} (2\pi f_0)^2}, \quad (1)$$

where SNR is the signal to noise ratio within the frequency band Δf of the signal, T is the signal duration and f_0 the signals fundamental frequency. For the windowed sine waves used in this study the signal bandwidth Δf is approximately $2/T$, i.e., $\Delta f T = 2$. For the white noise in these simulations $\text{SNR} = \text{SNR}_0 f_s / 2 / \Delta f$ where f_s is the sample rate and SNR_0 the total SNR over the full bandwidth of the signal.

3. Results

Figure 2 shows histograms of errors on timing measurements for five algorithms operating in the absence of noise. For simple cross correlation a high proportion (55%) of

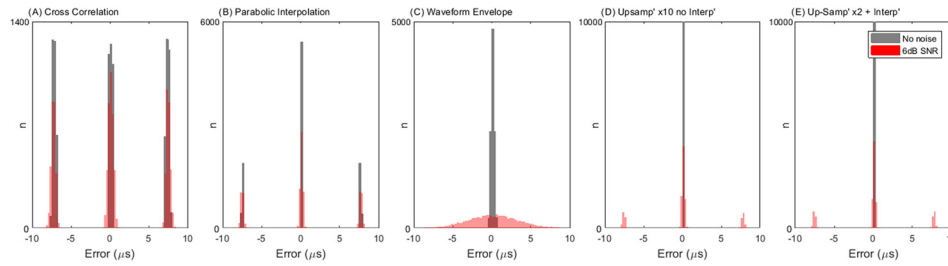


Fig. 2. (Color online) Histograms of residual errors on timing measurements for five timing algorithms in the presence and absence of noise. (A) Cross correlation with no interpolation. (B) Cross correlation with parabolic interpolation and peak selection. (C) Envelope tracing with parabolic interpolation and peak finding. (D) Up-sampling ($\times 10$) with no interpolation. (E) Up-sampling ($\times 2$) with parabolic interpolation and peak finding.

measurements contain gross errors indicating that the wrong cross correlation peak has been selected. When parabolic interpolation is used to help identify the correct peak, the number of gross errors reduces to 25% but the problem is not eliminated. Neither up-sampled data nor enveloped data have any gross errors in the absence of noise. When noise is added, a slight broadening of the peaks can be seen, indicating increased small errors. Small numbers of gross errors (13%) also start to appear even for the up-sampled data. There are no gross errors for the waveform envelope method, but overall in the presence of noise, there is a clear broadening of the range of values.

Figure 3 shows the relationship between errors using each method and SNR. At very poor SNR, no information at all is being obtained from the timing measurements, the errors being equal to random selection of measurements within the range of possible values ($\pm 100 \mu\text{s}$). The worst performing method is the simple cross correlation function. Errors are reduced slightly using parabolic interpolation, but even at high SNR incorrect peak selection still affects around 35% of the data, meaning that the RMS error does not reduce to below $5 \mu\text{s}$ (or 2.5 samples at 500 kHz sample rate). The two methods that are best able to provide accurate TOAD measurements are up-sampling the data by a factor 2 and using the waveform envelope in the cross correlation function. At SNRs between 0 and 8 dB, the envelope method shows marginally better performance which is due to the methods ability to more often find the correct correlation peak. At SNR above 10 dB, the $\times 2$ up-sampled data with parabolic interpolation has the best performance, with errors falling to below $0.1 \mu\text{s}$ at SNRs above 15 dB. At SNRs between 15 and 25 dB, the error is close to the expected Cramer Rao lower bound. As SNR increases however, the error ceases to decrease much below $0.01 \mu\text{s}$, or 1/200th of a sample.

Figure 4 shows TOAD measurements between two channels of data from a sequence of porpoise clicks recorded on a SoundTrap recorder. These clicks mostly had SNRs measured in the 100–150 kHz band of between 20 and 35 dB. Although the true TOAD is not known for these data, an animal swimming by the recorder is constrained by swim speed and so will produce TOAD measurements which lie on a smooth line. TOAD measurements using up-sampling lie on this expected smooth curve, TOADs using envelope tracing are scattered randomly about that curve, indicating higher errors and both methods that do not use up-sampling show gross errors consistent with incorrect correlation peak selection.

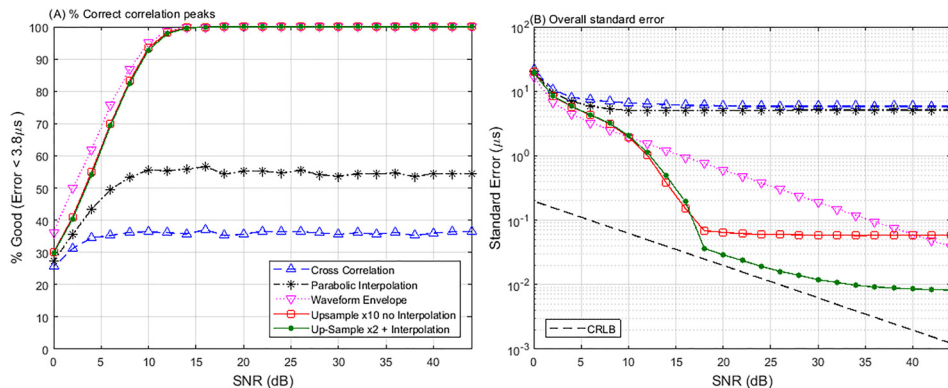


Fig. 3. (Color online) (A) The percentage of measurements which achieve an error $< 5 \mu\text{s}$, indicating correct correlation peak selection. (B) Timing errors as a function of SNR for different timing algorithms. The straight dashed line is the theoretical Cramer-Rao lower bound for signal waveform correlation.

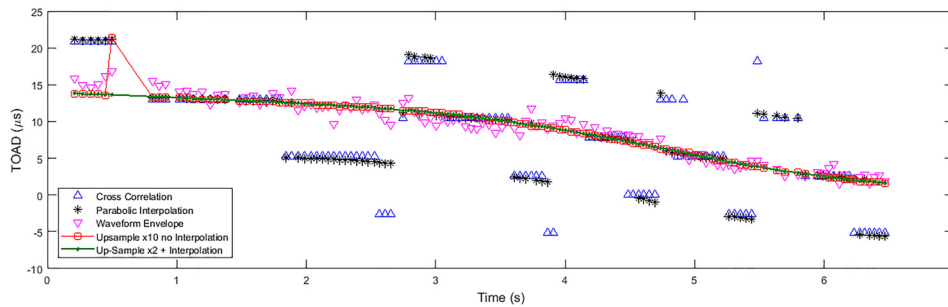


Fig. 4. (Color online) Time of arrival measurements for a 7 s long sequence of porpoise clicks using five different time delay algorithms.

4. Discussion

Even in the absence of noise, cross correlation of narrow band high frequency echolocation clicks is prone to high numbers of gross errors caused by incorrect peak selection in the cross correlation function. These errors will contribute much greater uncertainty to localizations derived from TOAD measurements than small errors of <0.5 samples obtained when the correct correlation peak is selected. Parabolic peak interpolation can help to both identify the correct peak in the correlation function and also improve the timing accuracy around that peak. However, for data sampled at only a few times the fundamental frequency of the data, this is not sufficient to correctly resolve the correct correlation peak, even in the absence of noise. Two methods that can significantly improve peak selection accuracy are up-sampling the data to at least 6 times the fundamental frequency and envelope tracing of the signal. Parabolic interpolation of data up-sampled by a factor 2 provided more accurate results than data up-sampled by a factor of 10 without interpolation. Without interpolation, accuracy was limited to the ± 0.5 samples discussed in Sec. 2.3, which for 500 kHz, $10\times$ up-sampled data is $0.06\ \mu\text{s}$. With parabolic interpolation, at increasing SNR accuracy did not improve much beyond $0.01\ \mu\text{s}$. This is due to the parabolic interpolation being an imperfect representation of a sinusoid. The effects of up-sampling have been further investigated with simulated data generated with varying sample rates: It was found that if the sample rate is around three times the fundamental frequency of the waveform, then further improvement can be gained by up-sampling by a factor 3 rather than 2, but that up-sampling cannot recover accurate time information when the sample rate is below 2.5 times the fundamental frequency for this type of signal.

While filtering was applied to the data in the examples shown here, its effect was no different to improving the signal to noise ratio by 7 dB. However, the simulated noise was spectrally flat and incoherent on the two data channels. Real ocean noise is generally more dominant at low frequencies (Urlick, 1975) and may also be more coherent on the two channels. Filtering is therefore likely to have a much more significant impact with real data than is shown here.

The accuracy of a bearing measurement derived from a TOAD measurement is related to the ratio of the time error measurement and the spacing between sensors. For two sensors, spaced a distance D apart, the angle of arrival θ can be calculated as

$$\theta = \cos^{-1}\left(\frac{ct}{D}\right), \tag{2}$$

where t is the TOAD measurement and c is the speed of sound in water. It then follows that the error $\delta\theta$ on θ is given by

$$\delta\theta = -\frac{c}{D \sin(\theta)} \delta t, \tag{3}$$

where δt is the error on the TOAD measurement t . Clearly, the magnitude of the error varies with angle, errors being large when the source is in line with the two sensors [as $\sin(\theta) \rightarrow 0$]. For angles $>45^\circ$ a 1° error can be obtained if the timing error is $<1\ \mu\text{s}$ with sensors spaced as close as 5 cm apart which is clearly attainable for clicks with an SNR above 12 dB.

The above algorithms have been implemented in the open source detection classification and localization software PAMGuard (Gillespie *et al.*, 2008) and can be applied to data being processed in real time or re-processed offline. Execution times varied for the different methods, simple correlation or correlation with parabolic

interpolation taking approximately 1 ms per TOAD measurement running on a 2.7 GHz Intel® i7 processor. The envelope and up-sampling by a factor 2 methods took only marginally longer, increasing CPU usage by only 4% and 5%, respectively. Up-sampling by a factor 10 took 6 ms per calculation making it the slowest method tested. It is often the case that multiple species are being detected and localized within the same dataset. The algorithm choice (particularly with regard to filtering) may therefore vary with the species being detected. The PAMGuard implementation therefore applies the timing algorithm after clicks have been identified to species in order that species specific timing choices may be applied.

Acknowledgments

Funding for this research was provided by the Scottish Government as part of the Marine Mammal Scientific Research Program MMSS/002/15 and from the Natural Environment Research Council, Grant No. NE/R014639/1. Development of a standardized marine mammal monitoring system for the tidal energy industry. Porpoise click data were collected under a DEFRA funded project ecm_54439 “Continued development and deployment of a passive acoustic tracking system on gill nets.” We are also grateful to two earlier reviewers of this manuscript for their useful input.

References and links

- Fisher, R. B., and Naidu, D. K. (1996). “A comparison of algorithms for subpixel peak detection,” in *Image Technology*, edited by J. L. C. Sanz (Springer, Berlin), pp. 385–404.
- Gasiot, M. (2004). “Improving FFT frequency measurement resolution by parabolic and Gaussian spectrum interpolation,” *AIP Conf. Proc.* **732**, 276–285.
- Gillespie, D., Gordon, J., Mchugh, R., McLaren, D., Mellinger, D., Redmond, P., Thode, A., Trinder, P., and Deng, X. Y. (2008). “PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans,” *J. Acoust. Soc. Am.* **125**, 2457.
- Kyhn, L. A., Jensen, F. H., Beedholm, K., Tougaard, J., Hansen, M., and Madsen, P. T. (2010). “Echolocation in sympatric Peale’s dolphins (*Lagenorhynchus australis*) and Commerson’s dolphins (*Cephalorhynchus commersonii*) producing narrow-band high-frequency clicks,” *J. Exp. Biol.* **213**, 1940–1949.
- Macaulay, J., Gordon, J., Gillespie, D., Malinka, C., and Northridge, S. (2017). “Passive acoustic methods for fine-scale tracking of harbour porpoises in tidal rapids,” *J. Acoust. Soc. Am.* **141**, 1120–1132.
- Sousa-Lima, R. S., Norris, T. F., Oswald, J. N., and Fernandes, D. P. (2013). “A review and inventory of fixed autonomous recorders for passive acoustic monitoring of marine mammals,” *Aquat. Mamm.* **39**, 23–53.
- Teilmann, J., Miller, L. A., Kirketerp, T., Kastelein, R. A., Madsen, P. T., Nielsen, B. K., and Au, W. W. (2002). “Characteristics of echolocation signals used by a harbour porpoise (*Phocoena phocoena*) in a target detection experiment,” *Aquat. Mamm.* **28**, 275–284.
- Urick, R. J. (1975). *Principles of Underwater Sound* (McGraw-Hill, New York).
- Weiss, A., and Weinstein, E. (1983). “Fundamental limitations in passive time delay estimation—Part I: Narrow-band systems,” *IEEE Trans. Acoust. Speech Signal Process.* **31**, 472–486.