

A generalized power-law detection algorithm for humpback whale vocalizations

Tyler A. Helble,^{a)} Glenn R. Lerley, and Gerald L. D'Spain

Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0701

Marie A. Roch

Department of Computer Science, San Diego State University, 5500 Campanile Drive, San Diego, California 92182-7720

John A. Hildebrand

Scripps Institution of Oceanography, University of California at San Diego, La Jolla, California 92093-0701

(Received 25 July 2011; revised 9 January 2012; accepted 17 January 2012)

Conventional detection of humpback vocalizations is often based on frequency summation of band-limited spectrograms under the assumption that energy (square of the Fourier amplitude) is the appropriate metric. Power-law detectors allow for a higher power of the Fourier amplitude, appropriate when the signal occupies a limited but unknown subset of these frequencies. Shipping noise is non-stationary and colored and problematic for many marine mammal detection algorithms. Modifications to the standard power-law form are introduced to minimize the effects of this noise. These same modifications also allow for a fixed detection threshold, applicable to broadly varying ocean acoustic environments. The detection algorithm is general enough to detect all types of humpback vocalizations. Tests presented in this paper show this algorithm matches human detection performance with an acceptably small probability of false alarms ($P_{FA} < 6\%$) for even the noisiest environments. The detector outperforms energy detection techniques, providing a probability of detection $P_D = 95\%$ for $P_{FA} < 5\%$ for three acoustic deployments, compared to $P_{FA} > 40\%$ for two energy-based techniques. The generalized power-law detector also can be used for basic parameter estimation and can be adapted for other types of transient sounds. © 2012 Acoustical Society of America. [DOI: 10.1121/1.3685790]

PACS number(s): 43.30.Sf, 43.60.Cg, 43.80.Ka [WWA]

Pages: 2682–2699

I. INTRODUCTION

Detecting humpback whale (*Megaptera novaeangliae*) vocalizations from acoustic records has proven to be difficult for automated detection algorithms. Humpback songs consist of a sequence of discrete sound elements, called units, that are separated by silence.¹ Both the units and their sequence evolve over time and cover a wide range of frequencies and durations.^{1,2} In addition, individual units may not repeat in a predictable manner, especially during non-song or broken song vocalizations, or in the presence of multiple singers with overlapping songs.^{1,2} Many types of marine mammal detection and classification techniques have been developed, using methods of spectrogram correlation,³ neural networks,⁴ hidden Markov models,^{5,6} and frequency contour tracking,⁷ among others. Depending on the species of marine mammal, noise condition, and type of vocalization, many of these methods have been shown to be effective in producing high probabilities of detection (P_D) with low probabilities of false alarm (P_{FA}). However, for humpback vocalizations, these techniques often provide low P_D if the P_{FA} is to remain adequately low. Abbot *et al.*⁸ used a kernel-based spectrogram correlation to identify the presence of humpback whales with extremely low P_{FA} . However, their approach requires 15 kernel matches within a 3 min window to trigger a detection. Therefore the

goal is not to detect every humpback unit but rather to predict the presence of song when enough predefined kernels are matched. Energy detection algorithms, readily available in acoustic analysis software such as ISHMAEL,⁹ XBAT,¹⁰ and PAM-GUARD¹¹ have proven effective for detecting all types of humpback call units. However, to avoid an exorbitant number of false detections, these methods generally require high signal-to-noise ratio (SNR): The hydrophones are in close proximity to the whales, and/or the shipping noise is low. Erbe and King¹² recently developed an entropy detector that can outperform energy detection methods for a variety of marine mammal vocalizations. However, this method is inadequate for detecting humpback vocalizations for data sets that contain considerable shipping noise. Therefore a need still exists for an automated detection capability in low SNR scenarios that is able to achieve low probability of false alarms, yet is general enough to achieve high probability of detection for all humpback units, including those with poorly defined spectral characteristics.

Nuttall introduced a general class of power-law detectors for a white noise environment.^{13,14} The energy method—based on the square of the Fourier amplitude—is a particular case, optimum when the signal occupies all the frequency bands over which energy summation occurs. However, in the case of narrowband transient signals that fall within a wide range of monitored frequencies (characteristic of humpback vocalizations), the optimal detector from Nuttall's work has a markedly higher power than the square. This paper builds on

^{a)}Author to whom correspondence should be addressed. Electronic mail: thelble@ucsd.edu

this insight but with suitable adaptation for the highly colored and variable noise environment characteristic of the Southern California Bight, notably containing interfering sounds from large transiting vessels. Unlike most commonly used detectors, the generalized power-law (GPL) detector introduced here uses detection threshold parameters that are robust enough not to require operator adjustments while reviewing deployments spanning months to years, with highly varying ocean noise conditions. Such a technique has the potential to significantly reduce operator analysis time for determining humpback presence/absence information as well as the capacity to determine basic call unit parameters, such as unit duration, that are normally time-prohibitive to obtain using manual techniques. The goal for this detector is to detect nearly all humanly audible humpback call units, allowing for occasional false detections in periods of heavy shipping. This detector is not designed to discriminate between transient biological signals that occur in overlapping spectral bands and of similar duration. However, the method has a limited capacity for classification; namely the ability to separate shipping noise from narrowband, transient signals. Therefore additional classification may be necessary if other acoustic sources meet the GPL detection criteria. Conversely, the GPL detector has proven to perform well for detecting other biological signals. In unpublished experiments, suitable selection of spectral analysis parameters has provided good results for detecting blue whale (*Balaenoptera musculus*) “D” calls, minke (*Balaenoptera acutorostrata*) “boings,” and killer whale (*Orcinus orca*) vocalizations in the Southern California Bight (blue and minke whales) and in the coastal waters of Washington State (killer whales).

This paper is divided into six parts: Sec. II describes commonly employed manual detection techniques, which guide the design constraints for an acceptable automated detector. Section III presents theoretical analysis for the GPL algorithm, highlighting the departures from the Nuttall form, which are motivated by these design constraints. Readers primarily interested in the application of the detector can move directly to Sec. IV, which discusses the particular application of the GPL algorithm to observational data, including the parameters chosen to best suit these data sets. Section V discusses the results of Monte Carlo simulations conducted to characterize the performance of the detector in comparison to: Nuttall’s original power-law processor, the Erbe and King entropy method, and two energy-based detection algorithms. These simulations provide detection error trade-off²⁴ (DET) curves for various humpback units, SNR, and noise conditions. In addition, results are given from simulations conducted to measure the performance of these algorithms against trained human analysts. Section VI quantifies the ability of the GPL algorithm to measure call duration parameters. Finally, Sec. VII presents the results from applying the GPL algorithm to 20 h of recordings from three different deployments where humpback units were previously marked by trained human analysts. These 60 h of acoustic data contain 21 037 individual humpback units occurring over a variety of ocean conditions and SNR. Although they perform poorly, the two energy detection algorithms are also included in this analysis because they are commonly used.

II. DETECTOR DESIGN CONSIDERATIONS

Detector design considerations were developed based on data sets collected by the Scripps Whale Acoustics Lab. However, similar detection requirements are representative of the needs of the marine mammal acoustics community in general. The data sets for detecting humpback vocalizations were recorded by high-frequency acoustic recording packages (HARP).¹⁵ These packages contain a hydrophone tethered above a seafloor-mounted instrument frame deployed in depths ranging from 200 to 1500 m, covering a wide geographic area in the southern California Bight, and record more or less continuously over all seasons. HARP data are used to study the range and distribution of a wide variety of vocalizing marine mammals. The first step is to identify marine mammal vocalizations in the data. Depending on the type of marine mammal, this process can be labor intensive. Humpback recordings are particularly difficult. Humpback units can be described as transient signals, the structure, strength, frequency, duration, and arrival time of which are unknown. Additionally, these vocalizations often occur in the same frequency bands that contain colored noise with additional contamination created by large transiting vessels. Depending on the distance of the passing ship, ship sounds can appear non-stationary over the same time scales as humpback units. The structure of the shipping noise is unknown but is often broadband. In practice, this complicated signal and noise environment often leads analysts to abandon automated detection entirely, relying on manual techniques for identifying vocalizations.

Various methodologies are used by the Whale Acoustics Lab to ensure consistent manual detection of marine mammal vocalizations. The TRITON software package¹⁶ was developed by the lab, providing the analyst with the ability to look at the time series and resulting spectrogram, with adjustable dynamic range, window lengths, filters, de-noising features, and audio playback. These manual detection techniques often find humpback units that are otherwise missed by standard automated detectors. While the ability to correctly mark the beginning and end time of each humpback unit is desirable, this step is time-prohibitive for longer data sets, and often only binary humpback presence/absence information is logged.

An acceptable automated humpback whale detector must be able to keep the probability of missed detections (P_{MD}) at or below the level of trained human analysts with a P_{FA} less than 6% in the noisiest environments. The amount of analyst review time required to separate humpback units from false detections depends upon both P_{FA} and the level of humpback vocalization activity. In practice, the 6% limit on P_{FA} necessitated 16 h of review for a 365 day continuously recorded deployment in the southern California Bight, containing greater than one million humpback units. A reliable fixed detection threshold which fits within these constraints is desired for the entire deployment. Additionally, the algorithm must run significantly faster than real-time and provide accurate humpback unit start times and end times.

III. THEORY

One approach for detecting signals with unknown location, structure, extent, and arbitrary strength is the power-law

processor. Using the likelihood ratio test, Nuttall derives the conditions for near-optimal performance of this processor in the presence of white noise, based on appropriate approximations.¹⁴ Nuttall's signal absent hypothesis (H_0) is equivalent to assuming that the short time Fourier transform (STFT) of the time series yields independent, identically distributed (iid) exponential random variables of unit norm. The signal present hypothesis (H_1) is that the STFT consists of two exponential populations. Wang and Willet¹⁷ represent these exponential populations as

$$H_0 : f(\mathbf{X}) = \prod_{k=1}^K \frac{1}{\lambda_0} e^{-|X_k|^2/\lambda_0},$$

$$H_1 : f(\mathbf{X}) = \prod_{k \in S} \frac{1}{\lambda_0} e^{-|X_k|^2/\lambda_0} \times \prod_{k \in S^c} \frac{1}{\lambda_1} e^{-|X_k|^2/\lambda_1} \quad (1)$$

where λ is the mean square amplitude, K is the total number of frequency bins, \mathbf{X} is the Fourier vector with components X_k , and S is the subset of size M , the number of frequency bins occupied by signal.

[Notation here and in succeeding sections is standard for probability theory:¹⁸ F is used to denote the cumulative distribution function (cdf) and f denotes the probability density function (pdf). In addition the upper case letters Y, Z denote general random variables and the lower case letters y, z are specific realizations of them. Owing to the particular needs of this paper, X is reserved for Fourier components. The upper case E indicates the expectation operator.] Application of the likelihood ratio test requires summing over all combinatorial possibilities in H_1 . For even moderate M , this step becomes infeasible. Hence, Nuttall develops various approximations to estimate a threshold for a power-law detection statistic of the form

$$T(\mathbf{X}) = \sum_{k=1}^K |X_k|^{2\nu}. \quad (2)$$

The variable ν is an adjustable exponent that can be optimized for a particular M . For the idealized case of white noise, Nuttall's work indicates a general purpose value of $\nu = 2.5$ when M is completely unknown. For a single snapshot in time, one can assume that for a humpback unit the number of signal bins M is much less than the total number of bins K , which favors $\nu > 2.5$. A summation of energy over all STFT bins is equivalent to $\nu = 1$, which is only optimal for $M = K$, and hence inappropriate here. Nonetheless, it is used extensively in readily available marine mammal detection software, and so its performance is noted throughout this paper.

A complication in the determination of an optimal ν is that most data sets contain shipping sounds in addition to the colored noise typical of the marine environment. A trade-off is created between values of ν that favor humpback vocalizations and larger values that better discriminate against broadband shipping sounds. No single choice of ν can be ideal for both purposes; however, a GPL detector can achieve a suitable compromise between these alternatives as well as a

fixed threshold in all noise environments. The definition of this detection problem is as follows:

$$H_0 : \begin{cases} n(t) & \text{or} \\ n(t) + s_1(t), \end{cases}$$

$$H_1 : \begin{cases} n(t) + s_2(t) & \text{or} \\ n(t) + s_1(t) + s_2(t) \end{cases} \quad (3)$$

where $n(t)$ is a time series generated from distant shipping and wind, which can be modeled as a Gaussian distributed stochastic process. Local shipping sounds created by a single nearby ship are represented by $s_1(t)$, which can be both non-stationary and contain intermittent coherent broadband structure in frequency. The quantity $s_2(t)$ is the humpback vocalization signal. Although not a contributing factor in the frequency bands of the data sets used in this work, any additional acoustic sources determined not to be humpback whales are also considered noise, and categorized as H_0 . Associated with these hypotheses is a formal optimization problem subject to nonlinear inequality constraints:

$$\min_{\Theta} P_{FA}(T^g(\mathbf{X}; \Theta)) \quad (4)$$

subject to

$$P(T^g(\mathbf{X}; \Theta) < \eta_{\text{thresh}} | H_1) = P_{MD} \leq P_{MD}^H,$$

$$P(T^g(\mathbf{X}; \Theta) > \eta_{\text{thresh}} | H_0) = P_{FA} \leq P_{FA}^{\text{max}} \quad (5)$$

where $T^g(\mathbf{X}; \Theta)$ is the generalized power-law detection statistic, η_{thresh} is the detector threshold value, P_{FA} is detector probability of false alarms, P_{FA}^{max} is the upper bound on false alarms (6%), P_{MD} is the detector probability of missed detection, P_{MD}^H is the human probability of missed detection, and Θ is model parameters.

Hereafter, the argument Θ will be dropped, its dependence implicit. Note that the superscript g distinguishes the GPL power-law detector from the Nuttall form.

To be considered an acceptable solution, a constant set of values for Θ , including η_{thresh} , is necessary. As in many other constrained optimization problems, the optimal solution is likely to be attained by an end-point minimum. A more traditional approach would be to permit detection on both $s_1(t)$ and $s_2(t)$, deferring discrimination to subsequent classification. While further classification is always possible, it turns out that this discrimination can be done largely at the detection stage if the power-law processor is suitably adapted. This goal is in the spirit of Wang and Willet,¹⁷ who developed a plug-in transient detector suitably adapted for a colored noise environment.

The characteristics described for $s_1(t)$ require examination of whitening, normalization, and broadband noise suppression. The non-stationary nature of $s_1(t)$ and the time clustered nature of $s_2(t)$ together motivate the choice of a conditional whitener insensitive to outliers. Similarly, while stationary noise motivates a simple estimator to produce the desired unit mean noise level, this normalization is less

appropriate for the varying noise environments of H_0 , where it is more important to bound the largest values generated by the test statistic. Last, broadband suppression requires unit normalization across frequency in addition to normalization within frequency.

Another consideration is discrimination based on temporal persistence of the test statistic. Provided ν is appropriately chosen, local shipping characteristically generates highly intermittent values of the test statistic while humpback vocalizations exhibit continuity in the test statistic over the typically longer duration of the call unit. An event is defined as a continuous sequence of test statistic values at least one of which exceeds a prescribed value η_{thresh} and which is delimited on each side by the first point for which the test statistic is at or below η_{noise} , a noise baseline. The expectation with this definition is that an event corresponds to a humpback call unit, and as such a minimum unit duration, τ_c , is a reasonable additional model parameter to incorporate into the detector (discussed in Sec. IV). Because the statistical distributions $H_{0,1}$ cannot be solved for analytically, η_{thresh} and η_{noise} are determined empirically with guidance from theory.

The proposed modification of the power-law statistic that incorporates these adaptations and also reflects the time dependence, j , can be written in its most general form as

$$T^g(\mathbf{X})_j = \sum_{k=1}^K a_{k,j}^{2\nu_1} b_{k,j}^{2\nu_2} \equiv \sum_{k=1}^K n_{k,j}, \quad (6)$$

$$a_{k,j} = \frac{||X_{k,j}||^\gamma - \mu_k}{\sqrt{\sum_{n=1}^K (||X_{n,j}||^\gamma - \mu_n)^2}}, \quad (7)$$

$$b_{k,j} = \frac{||X_{k,j}||^\gamma - \mu_k}{\sqrt{\sum_{m=1}^J (||X_{k,m}||^\gamma - \mu_k)^2}} \quad (8)$$

where \mathbf{X} now represents a Fourier matrix with J STFTs, j is the snapshot index ranging from 1 to J ; k is the frequency index ranging from 1 to K ; $\{a, b, n\}_{k,j}$ are elements in the matrices \mathbf{A} , \mathbf{B} , and \mathbf{N} , respectively, ν_1, ν_2, γ are adjustable exponents, μ_k is the conditional whitener, defined in the following text.

It is helpful to note that \mathbf{A} is a matrix the columns of which are of unit length. The normalization across frequency [Eq. (7)] enforces the desired broadband suppression. \mathbf{B} is a matrix the rows of which are of unit length, resulting from a normalization across time [Eq. (8)]. The average μ_k is defined by

$$\mu_k = \int_0^\infty zf_k(z)dz. \quad (9)$$

For the purpose of whitening, this is approximated by

$$\mu_k \approx \int_{F_k^{-1}(y_c)}^{F_k^{-1}(y_c+1/2)} zf_k(z)dz, \quad (10)$$

$$y_c = \min_{y \in [0,1/2]} [F_k^{-1}(y+1/2) - F_k^{-1}(y)]. \quad (11)$$

Equation (10) includes 50% of the distribution centered about the steepest part of the cdf, corresponding to the peak of the pdf. This form is termed ‘‘conditional’’ to reflect that the limits of integration are dynamically determined from the data rather than fixed as in Eq. (9). This formula is one of several possible implementations of a whitener the goal of which is to suppress one or more strong signals, such as the order-truncate average.¹⁹ Equation (10) is unbiased for f_k , a symmetric pdf, but is biased to the low side for the skewed distributions of interest here. The bias is not large however hence a more elaborate estimator of μ_k has not been explored. The integrals are cast in discrete form as follows. Let s_j denote the sorted values (from small to large) of $|X_{k,j}|$ over $j=1, \dots, J$ for a fixed k . Next find $j^* = \min_j (s_{j+J/2-1} - s_j)$. And finally

$$\mu_k = \frac{2^{j^*+J/2-1}}{J} \sum_{j=j^*} s_j.$$

The conditional restriction of the average to those points deemed in the noise level means that the numerators in Eqs. (7) and (8) using the μ_k in the preceding text are not exactly zero mean, although small.

Obtaining analytical expressions in the analysis of Eqs. (6) to (11) for $H_{0,1}$ is a difficult task. However, the case of white noise permits reasonable progress in characterizing the normalization and the whitener, which are explored in the following subsections. For white noise, only the sum $\nu_1 + \nu_2$ matters and hence can be replaced by a single exponent ν . For conditions other than white noise, the choices of γ, ν_1 , and ν_2 must be set individually, deviating from Nuttall’s one parameter form. For the optimization problem stated in Eqs. (4) and (5), values of $\gamma = 1, \nu_1 = 1$, and $\nu_2 = 2$ yielded about the minimal P_{FA} . These values were obtained with the guidance of theory presented in the following subsections and verified with Monte Carlo simulations and observational results. In the remainder of the paper, these are the values employed.

A. Statistics of unit normalization for white noise

To understand the importance of the normalized variables that enter into Eq. (6), consider the case of white noise. In this section, the focus is on normalization and hence μ_k is set to zero in Eq. (6). To represent the associated Fourier coefficients X_k let

$$X_k = \frac{1}{\sqrt{2}}(\Re(X_k) + i\Im(X_k)) \quad (12)$$

where real and imaginary parts are each independent and identically distributed normal random variables of zero mean and unit variance. With this normalization, $|X_k|$ has a Rayleigh distribution, $E(|X_k|) = \sqrt{\pi}/2$, and $E(|X_k|^2) = 1$, independent of frequency.

First consider the statistics of $a_{k,j}^2$ alone, hence define the random variable Y by

$$Y = \frac{|X_k|^2}{\sum_{n=1}^K |X_n|^2}, \quad (13)$$

where K is the number of Fourier frequency bins in the retained band. The matrix column index is omitted for the moment. The pdf for Y , $f_Y(y)$, is now sought. Because the sum in the denominator includes the index k , it is not independent of the numerator. Accordingly it is useful to look instead at the reciprocal, which is denoted as $1 + Z$ where Z is then given by

$$Z = \frac{\sum_{n=1}^{K'} |X_n|^2}{|X_k|^2}. \quad (14)$$

and the prime on the sum denotes the restriction $n \neq k$. From this starting point, standard statistical arguments lead to the conclusion that Y has the exact pdf

$$f_Y(y) = (K - 1)(1 - y)^{K-2}. \quad (15)$$

(See the appendix for details. In practice a Hamming window is used with the STFT and so this result does not strictly apply. The practical differences in the distributions obtained with a window compared to those in the preceding text are slight however.) From Eq. (15), it follows that $E(y) = 1/K$. Note that, also as expected from the normalized form, y is necessarily limited in range to $[0,1]$. This reflects the stated preference of bounding the test statistic in lieu of enforcing a unit norm of the noise as found in most implementations of the power-law processor. In the present case of white noise, the distinction is trivial, but such a bound remains in force even for the complex environments of $H_{0,1}$.

Equation (15) is well approximated by the exponential form $(K - 1) \exp(-(K - 2)y)$ provided $\log(1 - y) \approx -y$. The result is not, however, exactly normalized. To form a suitable pdf, it is appropriate to modify this expression to

$$f_Y(y) \sim (K - 2)e^{-(K-2)y}, \quad (16)$$

which has the proper unit area. A measure of the approximation error is seen in the modified mean, $E(y) = 1/(K - 2)$, which agrees with the exact result to only leading order in K .

While Eq. (15) correctly incorporates the fact that y can never exceed unity, a consequence of the expansion is that Eq. (16) has an exponentially small tail extending to infinity.

As shown in the Appendix, for even the simplest product of **A** and **B**, the statistics cannot be found in closed form. However, observe that if the denominator in Eq. (13) is replaced by its mean value of K , then the pdf for Y becomes simply a rescaled version of the numerator, namely $K \exp(-Ky)$. This last result, while not formally asymptotic to Eq. (16), is nonetheless a useful approximation for large K , and hence in subsequent sections when values are referred back to Eqs. (6) to (8), all normalizations are replaced by their mean values.

B. Unnormalized statistics for white noise only with mean removal

It is important to characterize the role of nonzero μ_k . The particular frequency is irrelevant hence the subscript k is

dropped in this subsection and subsection C. For this purpose, it is simplest to consider the unnormalized sum

$$Y = \sum_{n=1}^N ||X_n| - \mu|^p \quad (17)$$

where, with reference to Eq. (6), $p = 2\nu_1 + 2\nu_2$, leaving the summation index N general. In later plots, $p = [2, 6, \infty]$ are considered. The first of these, $p = 2$, addresses statistics of the denominators in Eqs. (7) and (8), the last two cover the numerators of interest. The value of p can be regarded in visual terms as a contrast setting; small p corresponds to low contrast, large p corresponds to high contrast, where ν_1 controls vertical contrast and ν_2 controls horizontal contrast through the relative weighting of the normalization (denominator) terms in Eqs. (7) and (8).

At certain points in this and the succeeding subsection, it is useful to form the related quantity

$$\left(\sum_{n=1}^N ||X_n| - \mu|^p \right)^{1/p}, \quad (18)$$

the classical L^p norm in \mathbb{R}^N to facilitate comparison of differing values of p . The limit of large p in this latter form yields the minimax, or infinity, norm, which singles out the largest single entry in the k th column. Using a measure with all its support concentrated at one point is probably not a good idea because humpback units commonly include very sharp upsweeps and downsweeps as well as units with a number of harmonics of similar amplitudes. Additionally, if p is too large, temporal persistence of the test statistic is lost and discrimination between shipping and transients such as humpback units is compromised. As previously indicated, the optimal constrained solution of Eqs. (4) and (5) is achieved in the neighborhood of $(\nu_1 = 1, \nu_2 = 2)$ or equivalently $p = 6$.

Now $|X_n|$ is Rayleigh distributed with, as noted before, a mean of $\sqrt{\pi}/2$. Defining the random variable

$$Z = ||X_n| - \mu|^p, \quad (19)$$

the associated pdf follows by a change of independent variable (see Appendix). The mean, $\mu_Z^{(p)}$, and standard deviation, $\sigma_Z^{(p)}$, of Z can be calculated but the expressions become unwieldy so the exact result is given only for $p = 2$ in Table I. The superscript (p) denotes the dependence on the exponent in Eq. (17). The salient features are: the value of moments grows exponentially with p and rate of exponential growth itself increases rapidly with the order of the moment. Hence the numerator and denominator in Eq. (6) do not approach the prediction of the central limit theorem at the same rate.

Evaluation of the N -fold convolution integral that represents the pdf for the sums in numerator and denominator leads to approximation in terms of the moment expansion of the characteristic function, of which the leading contribution is given exactly by the central limit theorem. On this basis it is expected that Eq. (17) is well approximated as

TABLE I. Distribution of moments for Eq. (17).

P	$\mu_Z^{(p)}$	$(\sigma_Z^{(p)})^2$	$\rho_Z^{(p)}$
2	$1 - \pi/4$	$1 + \pi/2 - \pi^2/4$	$2 + 15\pi/8 - \pi^3/4$
4	0.1494	0.4842	0.6481×10^1
5	0.1663	0.1613×10^1	0.7703×10^2
6	0.2154	0.6654×10^1	0.1257×10^4
22	0.7885×10^5	0.1922×10^{18}	0.2279×10^{33}

$$Y \approx \mu_Z^{(p)}N + \sigma_Z^{(p)}N^{1/2}z_d \quad (20)$$

for sufficiently large N , where z_d is a normally distributed random variable of zero mean and unit variance. However, it remains to be shown whether or not the asymptotic normal form is in fact an accurate approximation of the actual distribution for parameter values that are typical in application.

The first correction to the Gaussian pdf is the skewness, given by

$$c_3 = \int_{-\infty}^{\infty} Z_d^3 f_{Z_d} dZ_d = \frac{\rho_Z^{(p)}}{6\sqrt{2N}\pi(\sigma_Z^{(p)})^3},$$

and $\rho_Z^{(p)} = E(|Z|^3)$. Scaling the random variable by $\sqrt{2N}\sigma_Z^{(p)}$ to express it in terms of z_d , the corrected pdf assumes the form

$$f_Y \sim e^{-z_d^2/2} (1 + c_3 z_d (z_d^2 - 3)).$$

This is a good approximation provided

$$|z_d| \ll \sqrt[3]{6/\rho_Z^{(p)}N^{1/6}\sigma_Z^{(p)}}.$$

For $p=2$, i.e., the denominator in Eq. (6), this results in $c_3=0.0150$ valid for $|z_d| \ll 3$ while for the numerator with $p=6$, the skewness is nearly twenty times larger at $c_3=0.2644$ and consequently the expansion holds for $|z_d| \ll 1$, i.e., only the immediate vicinity of the peak of the pdf. Characterization of the tail of the distribution is given in the following text.

Figure 1 shows computed pdfs for the L^p norm in Eq. (18) for $p=2, 6, \infty$ along with the Gaussian pdf for comparison. It is seen that $p=2$ lies close to the normal distribution while $p=6$ is reasonably close to the infinity norm pdf. This bears directly on the analysis in the final theory subsection.

Turning briefly to the tails of these distributions, see Fig. 2 where $\log(1 - F_Y)$ is plotted. The parabolic curves in each panel reflect the quadratic controlling factor in the asymptotic expansion of the error function. This factor deviates significantly from the curve for $p=6$; the controlling factor in the correct cdf is weaker than linear. How much weaker is made clear by switching from a global representation to a local approximation, namely

$$\log(1 - F_Y) \sim -\sqrt[3]{N} \left(\sqrt{\pi}/2 + y^{1/6} \right)^2 + O(\log y). \quad (21)$$

Coefficients of the log and higher order corrections would derive from asymptotic matching. In lieu of that, here only

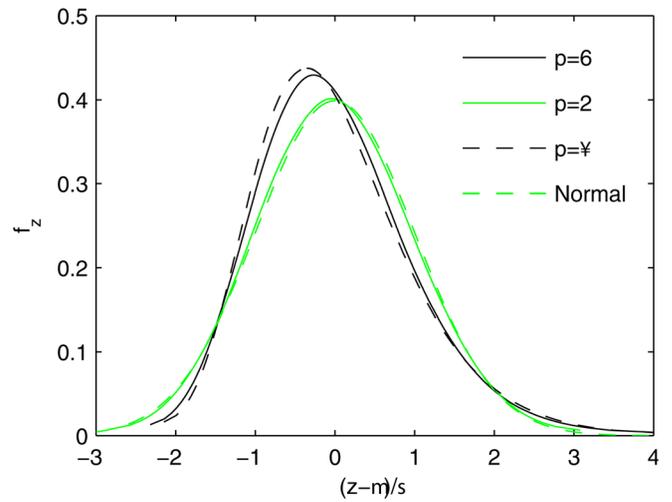


FIG. 1. (Color online) Computed pdfs for the L^p norm in Eq. (18) for $p=2, 6, \infty$ along with a Gaussian.

the first term is used along with a numerically determined constant offset.

The results in the preceding text individually characterize the numerator and denominator of Eq. (6). Because the

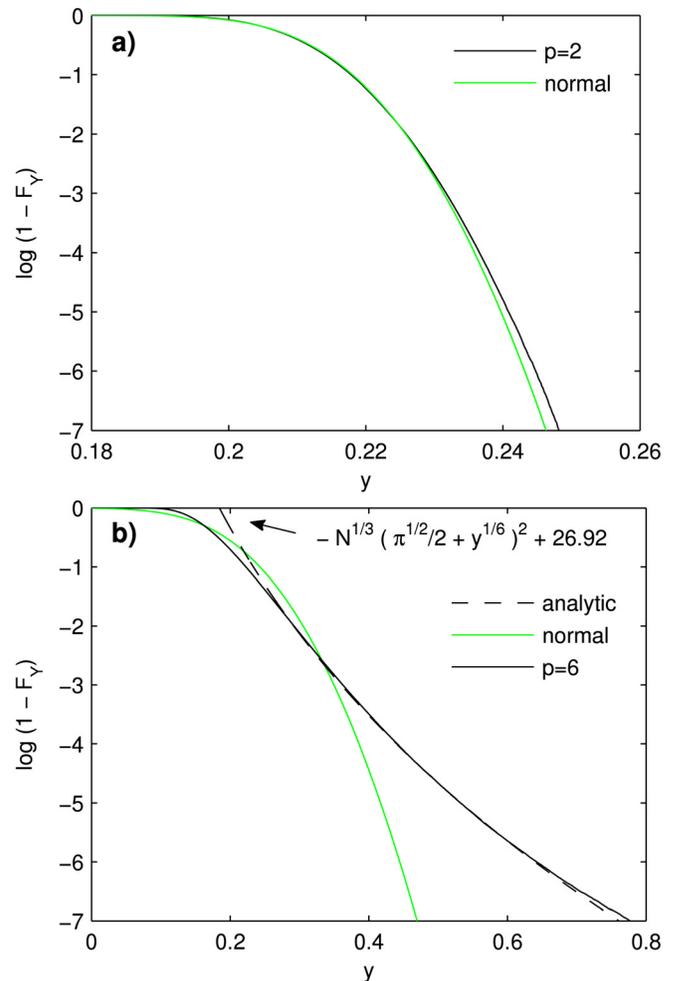


FIG. 2. (Color online) A comparison of numerical and analytic forms for the cdf of Eq. (17) for a) $p=2$ and b) $p=6$, emphasizing the tail of the distribution.

terms in the denominator have large mean with small relative variance, as previously noted in Sec. III A, little error is incurred by replacing them with their mean value. It is really the numerator alone that controls the distribution of $T^g(\mathbf{X})$. For a normalized detector based strictly on energy ($p=2$), no such partition is possible; the numerator and denominator scale comparably. This similarity of scaling is the basic cause of poor discrimination between shipping and humpback vocalizations for energy detectors.

The zeroth moment of the distribution is accurately estimated from the entries in Table I even though there is a long tail to the right, hence the average test statistic for H_0 is

$$\bar{T}^g(\mathbf{X}) \approx \frac{\mu_Z^{(p)}}{J^{p/2-1}(\mu_Z^{(2)})^{p/2}}, \quad (22)$$

independent of K . For $J=1460$, and $p=6$, this works out to a prediction of $\bar{T}^g(X) = 1.0223 \times 10^{-5}$. Simulations using Eq. (6) and the conditional whitener given in Eqs. (10) and (11) gives an average of 1.29×10^{-5} . In spite of real data leading to additional complications such as (1) overlap of successive spectra, (2) dependence of the μ_k on frequency, (3) nonstationarity of shipping noise, and (4) sensor self-noise (discussed in Sec. IV), it is notable that the operational noise threshold for use with HARP data is set at $\eta_{\text{noise}} = 2.07 \times 10^{-5}$, just a factor of two larger than the value from Eq. (22). Recall the purpose of η_{noise} is to delimit the beginning time and end time of a particular humpback unit. Therefore the final value was chosen to optimize the accuracy of this process, as described further in Sec. VI.

In lieu of a more elaborate model to incorporate the frequency dependence of μ_k , representative distributions are shown of $T^g(\mathbf{X})$ from recorded wind-driven noise, distant shipping, and local shipping data (discussed at greater length as cases 1,2,3 respectively in Sec. V) in comparison with the white noise result. In Fig. 3, a slightly different format for the tail of the distribution is used to bypass issues relating to a

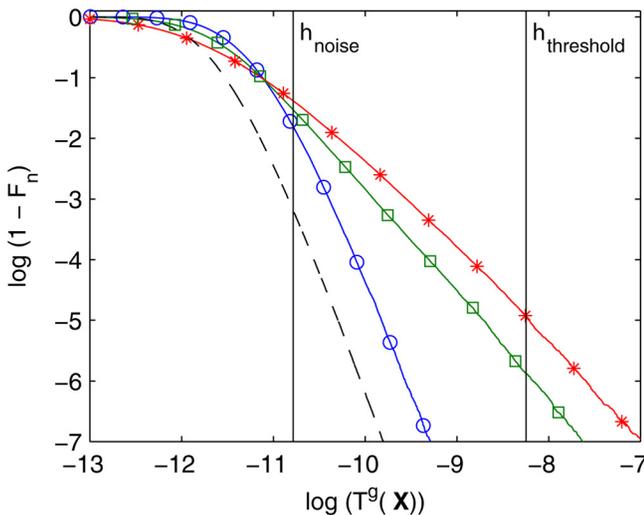


FIG. 3. (Color online) Comparison of the tails of the cdfs for local shipping (asterisk), distant shipping (open square), and wind-driven (open circle) noise conditions versus ideal white noise (dashed).

varying mean, μ_k , so the abscissa is now $\log(T^g(\mathbf{X}))$. Note how the tail of the wind-driven noise environment matches the ideal white noise result up to within a translation of about 0.5, which corresponds to a simple multiplicative rescaling of $T^g(\mathbf{X})$. The distributions of distant and local shipping, by contrast, decay more slowly although even for the latter on average a fraction of only about $\exp(-5)$ sample points per 75 s interval will exceed the indicated threshold. Whether these sample points produce an event detection is subject to the event duration requirement. Such persistent events come about not by a chance confluence of independent random spikes, which is quite rare, but from a spectral feature that does not fall to η_{noise} quickly enough to either side of the peak. How often that happens requires a more detailed model of shipping noise than is suitable to pursue here. A principal cause for excessively slow decay of the tail in Fig. 3 is failure of the whitener. During intervals of high level shipping, a prominent modulation of the spectrogram from ship propeller noise of a 10- to 20-s period typically occurs. In this case, the use of a constant μ_k at each frequency over a time window of 75 s leaves a significant residual sinusoidal modulation.

C. Signal plus noise

To understand the response of GPL in the simplest setting, the normalization can be omitted. Recall that its purpose is to allow fixed values for η_{noise} and η_{thresh} in $H_{0,1}$. With white noise of fixed variance, this normalization is unnecessary. It is helpful here also to use the standard L^p form

$$\tilde{T}^g(\mathbf{X})_j^{(p)} = \left[\sum_{k=1}^K |X_{k,j} - \mu|^p \right]^{1/p}. \quad (23)$$

The tilde denotes the absence of normalization in the remainder of this subsection. The main issue is the statistics of an isolated snapshot. The correlation of $\tilde{T}^g(X)_j^{(p)}$ with adjacent values $\tilde{T}^g(X)_{j\pm 1}^{(p)}$ arising from overlap of successive STFT windows is hence neglected here. While characterizing the pdf for $\tilde{T}^g(X)^{(p)}$ in analytic form is not easy for intermediate p , the limiting case of the infinity norm is relatively accessible. Moreover in Fig. 1, which shows the noise pdf for Eq. (23), the earlier noted similarity of results for $p = \infty$ and $p=6$ suggests that qualitative aspects of the analysis in the following text can be also expected to apply to the latter value of p .

For $p \rightarrow \infty$, Eq. (23) simplifies to

$$\tilde{T}^g(\mathbf{X})_j^{(\infty)} = \max_k |X_{k,j} - \mu|, \quad (24)$$

that is, the value assigned to \tilde{T} for time interval j is the single largest value in the k th column of the whitened amplitude matrix. As an idealized model of this process, the signal is assumed to be a sine wave of amplitude s that lasts exactly one snapshot, superimposed on white noise. Denote the index of its frequency as k' . (The actual value is irrelevant in what follows.) What matters is that the maximum in Eq. (24) is taken over K values in the frequency domain. One of these values contains the signal plus noise; the remaining $K - 1$

contain only noise. For this detection scheme to be reliable, the signal must be large enough that the corresponding value of $||X_{k',j}| - \mu|$ exceeds the likely extremal value over the remaining $K - 1$ realizations of pure noise.

The cdf for the case of pure noise is given by

$$\tilde{F}_n(z; K - 1) = \left(1 - \exp(-(z + \mu)^2)\right)^{K-1} \quad z > \mu. \quad (25)$$

For large K , the contribution in the range $z < \mu$ is exponentially small and may be neglected. The pdf for $||X_{k',j}| - \mu|$ is

$$\tilde{f}_s(z) = 2(z + \mu) \exp(-s^2 - (z + \mu)^2) \times \mathcal{I}_0(2s(z + \mu)), \quad z > \mu, \quad (26)$$

where \mathcal{I}_0 is the modified Bessel function of zeroth order. [For $0 \leq z \leq \mu$, the pdf is $\tilde{f}_s(z) + \tilde{f}_s(-z)$.] The accompanying cdf, $\tilde{F}_s(z)$, cannot be expressed in terms of known functions, however, its asymptotic and series expansions for large and small s , respectively, can both be found.

In terms of these quantities, the pdf for the random variable $z = \tilde{T}^g(X)$ summed over all frequencies including k' is given by

$$\tilde{f}_{GPL}^{(\infty)}(z) \sim \tilde{f}_s(z)\tilde{F}_n(z; K - 1) + \tilde{f}_n(z; K - 1)\tilde{F}_s(z), \quad (27)$$

with $K - 1$ equal to the total number of frequencies *not* counting that of the signal. From this construction, it follows automatically that $\int_0^\infty \tilde{f}_{GPL}^{(\infty)} dz = 1$. For large s and K , Eq. (27) has the simple leading order asymptotic expansion

$$\tilde{f}_{GPL}^{(\infty)}(z) \sim \sqrt{\frac{z + \mu}{\pi s}} e^{-(z + \mu - s)^2}, \quad (28)$$

which is an excellent approximation for $s \geq 4$.

From the derivative of Eq. (25), the pdf of noise for $\tilde{f}_{GPL}^{(\infty)}$ reaches a maximum at $z \sim \sqrt{\log(K - 1)} - \mu$. The predicted separation of the peaks of signal plus noise and noise only pdfs is thus $s - \sqrt{\log(K - 1)}$. Pressing Eq. (28) somewhat beyond its formal range of applicability in this last result suggests for $K = 339$ that $s > 2.4$ is required for a signal to begin to emerge from the background. This predicted separation is qualitatively corroborated in Fig. 4(a).

The case for the energy sum is given by Eq. (2) with $\nu = 1$. The sum of K noise terms has a cdf of $\Gamma(K, z)$. The pdf is well approximated by a normal distribution for the values of K considered here. The pdf for the signal follows from substituting $\mu = 0$ in Eq. (26) in the preceding text and then making a variable change to reflect the choice of energy rather than amplitude as the independent variable. Hence

$$f_s(z) = \exp(-s^2 - z)\mathcal{I}_0(2s\sqrt{z}). \quad (29)$$

The equivalent of Eq. (27) is then given by the convolution

$$f_E(z) = \frac{1}{\Gamma(K)} \int_0^z (z - x)^{K-1} e^{x-z} f_s(x) dx. \quad (30)$$

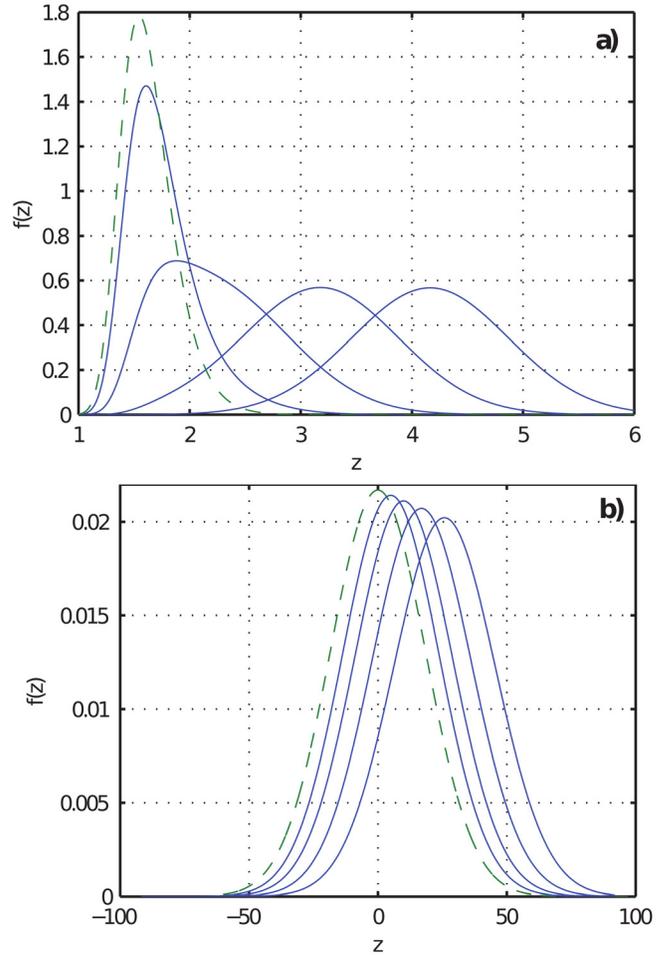


FIG. 4. (Color online) Pdfs for (a) $\tilde{f}_{GPL}^{(\infty)}$, (b) f_E for signal amplitudes of 0 (dashed) and 2, 3, 4, 5 (solid) from left to right in each plot.

This integral also cannot be found in closed form, but only approximated in various limits.

The displacement of the peak of f_E relative to the peak of the noise pdf at K is found to satisfy the approximate relation

$$4s^4 + (K - 1)(s^2 + z) = 2s^2 \frac{(2s^2 + K - 1)^{3/2}}{\sqrt{K - 1 + 2z}}, \quad (31)$$

which is equivalent to a cubic polynomial and has a K -independent exact root of $z = s^2$, as can be seen by inspection.

The plots in Fig. 4 show $\tilde{f}_{GPL}^{(\infty)}$ and f_E for signal amplitudes of $s = [0, 2, 3, 4, 5]$ (for, again, an rms noise amplitude of $\mu = \sqrt{\pi}/2$ per frequency and $K = 339$). Figure 4 suggests that it takes about a 5 dB dynamic range for GPL to go from essentially no detection to nearly perfect detection. Taking $s = 4$ to define a suitable threshold for detection, it is useful for orientation to convert this choice of s into an associated (normalized) value of η_{thresh} for $p = 6$. The denominator of $T^g(\mathbf{X})$ is estimated as previously in Eq. (22). For the numerator, it suffices to compute $\int_0^\infty z^6 \tilde{f}_s(z) dz$ with \tilde{f}_s as given in Eq. (26). The result is $\eta_{\text{thresh}} = 2.66 \times 10^{-4}$, virtually the exact value used in practice.

No algorithm based on $\nu = 1$ can compete with this performance; the linear separation of signal and noise with GPL

is complete before the quadratic separation of the energy method begins to be effective. A formal measure of SNR statistics is the deflection ratio, defined as

$$d = \frac{|\mu_{s+n} - \mu_n|}{\sqrt{\sigma_{s+n}^2 + \sigma_n^2}}. \quad (32)$$

Asymptotic expansions for the means are tedious, but for large K , the distinction between the mean values and the peaks of the corresponding pdfs is slight. Accordingly the latter are used instead, yielding

$$d_{GPL} \approx \frac{\sqrt{2}(s - \sqrt{\log(K-1)})}{1 + 1/(2\log(K-1))} \quad \text{and} \quad d_E \approx \frac{s^2}{\sqrt{2K}}. \quad (33)$$

The first of these reaches unit deflection ratio at $s = 3.2$, the second not until $s = 21.9$. Computed values of deflection ratio as defined in Eq. (32) based on statistics from simulations were compared against the analytical simplification for d_{GPL} in Eq. (33). Close agreement was found for $s > 4$, consistent with the approximation in Eq. (28) used to obtain d_{GPL} above. The computed values from simulation also corroborated a precise evaluation of Eq. (32) based on Gaussian quadrature with the exact pdf given in Eq. (27). Last, simulation confirms that $d_{GPL}(s)$ for $p = 6$ differs minimally from that for $p = \infty$ with an asymptotic slope reduced by only about 8%, thus discrimination for the ideal signal considered here is only slightly degraded by fixing $p = 6$ in place of the infinity norm, as anticipated.

Needless to say, real signals are not confined to a single frequency and the noise is neither white nor stationary. For these reasons, a more robust detector is required but one that nonetheless approximates this sifting property of the L^∞ norm. The choice of $p = 6$ ($\nu_1 = 1, \nu_2 = 2$) is a good compromise.

D. Summary

It is not hard to see why GPL (or any other optimized power-law processor) is good at practical noise rejection: An overwhelming fraction of the final sample points $\{T^s(\mathbf{X})\}$ is tightly clustered near $\bar{T}^s(X)$. These points, which lie below η_{noise} , automatically define the snapshots at which events begin and end. Their ubiquity ensures that although common noise sources (and ships particularly) do generate occasional spikes above threshold, the majority of the latter are subsequently discarded because their duration is nearly always less than the minimum unit duration subsequently imposed. More broadly, defining event duration is problematic for energy detection schemes both because no clean separation of signal and noise exists (equivalently the pdfs have excessive overlap) and because of the need to define an empirical *adaptive* threshold in contrast with the fixed value used in GPL.

What has been shown in the preceding subsections is that the modifications of normalization and whitening achieve white noise results comparable to those of Eq. (2). Analytical evaluation of these modifications in application to $H_{0,1}$ is not feasible. Rather, the evaluation is carried out in succeeding sections by means of both simulation and application to real data sets. It is shown that these modifications are necessary

for an acceptable solution to the constrained optimization problem in Eqs. (4) and (5) using real ocean acoustic data and cannot be achieved with the power-law processor in Eq. (2).

IV. SPECIFIC CONSIDERATIONS FOR GPL ALGORITHM USED ON HARP DATA FOR HUMPBACK DETECTION

HARP data are recorded in either continuous or duty cycled format with a sampling frequency of 200 kHz. For the results presented in this paper, data were processed in 75 s blocks, a time segment that was convenient for the duty cycle used in the HARP deployments. The time series is then low-pass filtered and decimated to a 10 kHz sampling rate. An STFT of length 2048 points is used with a 75% overlap and a Hamming window function, which corresponds to 4.9 Hz per frequency bin, 0.05 s per snapshot, and a total number of snapshots, J , equal to 1460. These parameters were found most effective for the majority of humpback vocalizations. The shortest call units could benefit from a shorter STFT length at the expense of a decrease in spectral resolution. No improvements in detection are realized for overlaps greater than 75%, therefore the overlap is fixed at 75% to avoid additional processing time. The output from the STFT is band-limited to a frequency range of 150–1800 Hz, and the number of frequency bins, K , is then 339. While humpback vocalizations can be recorded well above 1800 Hz and slightly below 150 Hz, sufficient energy for such units exists between these frequencies for good humpback detection performance.

The HARP data contain self-noise from the disk recording process. Therefore, a pattern matching algorithm based on singular value decomposition is used to remove short duration, broadband spectral features that coincide with the beginning and end of write-to-disk events. Additionally, the disk-write process produces narrowband, long duration (on the order of 10 s) noise contamination. While this narrowband noise is not problematic for higher order power-law processors, it does pose a problem for the energy-based detection methods (discussed in the following sections). Therefore for energy detection only, a second algorithm is deployed that searches for the five strongest frequencies containing these narrowband features and removes these bands in the spectrogram. For both the energy methods and GPL, $|\mathbf{X}|$ as defined in Eqs. (7) and (8) is whitened following the discretized version of Eqs. (10) and (11), defining $|\widehat{\mathbf{X}}| = \|\mathbf{X}_k| - \mu_k|$.

Threshold values were guided by both the theoretical calculations and the nonlinear inequality constraints discussed in Sec. III. Initially η_{thresh} was adjusted to match the performance of a trained human analyst. The theory in Sec. III provides an *ex post facto* analytical basis for this as a formal problem in separation of signal and noise. The simple choice of $s = 4$ gives a predicted η_{thresh} that lies fortuitously close to the chosen value.

It was found that values of $\eta_{\text{noise}} = 2.07 \times 10^{-5}$ and $\eta_{\text{thresh}} = 2.62 \times 10^{-4}$ satisfied these constraints while keeping $P_{FA} < P_{FA}^{\text{max}}$ in the heaviest shipping environments. The detection test statistics for each time step j are evaluated according to Eqs. (6) to (8) as earlier noted using $\gamma = 1$,

$\nu_1 = 1$, and $\nu_2 = 2$. Other values of γ , ν_1 , and ν_2 may be appropriate for other marine mammal vocalizations and/or noise conditions.

Using a normalized detection approach allows the user to set a fixed detection threshold, η_{thresh} , that works well over varying ocean conditions. However, during periods when the intercall interval between humpback units is short, the normalization approach reduces values of $T^s(\mathbf{X})$ for repeated units with shallow spectral slope, at times to values below η_{thresh} . Therefore an iterative method is used in an attempt to adjust $|\hat{\mathbf{X}}|$ so that $T^s(\mathbf{X})$ gives similar values for a particular call unit, regardless of call activity. First a preprocessing step is done: T^s is computed from $|\hat{\mathbf{X}}|$. A submatrix $|\hat{\mathbf{X}}|^s$ is formed containing all columns of $|\hat{\mathbf{X}}|$ for which the corresponding $T^s < \eta_{\text{noise}}$. Next T^s is recomputed from $|\hat{\mathbf{X}}|^s$ with J adjusted to the size of the submatrix. All columns of $|\hat{\mathbf{X}}|^s$ for which $T^s > \eta_{\text{thresh}}$ are removed. Iteration then proceeds as follows:

T^s is computed from $|\hat{\mathbf{X}}|$. The detection with the highest value of T^s that exceeds threshold is recorded, its duration n fixed by the nearest neighbor to either side for which $T^s < \eta_{\text{noise}}$. Next the n columns in $|\hat{\mathbf{X}}|$ corresponding to this event are replaced by n columns of $|\hat{\mathbf{X}}|^s$ chosen at random. The process is repeated until no values of T^s exceed η_{thresh} .

In rare cases where the unit is repeated heavily, the normalization that reduces shipping noise also reduces the contribution of the calls to the test statistic. In such cases, the statistic may be below the detection threshold. Alternative techniques for normalization have shown promise.

It is possible to further reduce the effects of shipping noise in the data using a minimum unit duration requirement as described in the following. After all events in the 75 s section of data have been determined, those events with a common terminus are merged into a single event. After qualifying events are merged, each event must exceed the minimum call duration requirement, τ_c , of 0.35 s. The modified detector output $T^{s*}(\mathbf{X})$ contains the values of $T^s(\mathbf{X})$ with detector values replaced by zero for events that do not meet these duration requirements. The formal optimization problems in Eqs. (4) and (5) should thus be changed so that $T^s(\mathbf{X})$ is replaced with $T^{s*}(\mathbf{X})$, and the model parameters contained in Θ are augmented to include $[\eta_{\text{thresh}}, \eta_{\text{noise}}, \tau_c]$. For an overlap of 75%, a minimum call unit duration of 0.35 s corresponds to seven snapshots. The event duration, τ , is recorded for each detection. Shipping noise can sometimes produce high values of $T^s(\mathbf{X})$ albeit short in duration. Most of these events are shorter than τ_c . Using energy techniques, detections from shipping events and humpback units occur on similar time scales, and so this method of discrimination cannot be utilized. For comparison purposes, the performance of $T^s(\mathbf{X})$ and $T^{s*}(\mathbf{X})$ are discussed in the following sections.

Because the event duration is computed from Fourier components rather than the original time series, STFT length and window overlap define the terminal points of the event.^{20,21} For example, due to the 75% overlap, energy occurring entirely within the snapshot j can influence the test statistic from $X_{k,j-3}$ to $X_{k,j+3}$. This overlap can hence permit detection of events slightly shorter than τ_c ; this is useful in

the case of detecting shorter humpback units but can also increase false detection from shipping noise.

An example of the GPL process can be seen in Fig. 5, the corresponding time series of which was created by adding a HARP recording containing strong shipping noise to a filtered HARP recording of humpback units (details discussed in Sec. V and shown in Fig. 6). Visual representations of \mathbf{X} , $|\hat{\mathbf{X}}|$, and \mathbf{N} for 30 s of data are shown in Figs. 5(a), 5(c), and 5(e). The incoherent sum over frequency for these matrices as a function of time are shown in Figs. 5(b), 5(d), and 5(f), where Fig. 5(b) represents the energy sum, Fig. 5(d) represents the whitened energy sum, and Fig. 5(f) shows the values of $T^s(\mathbf{X})$. In Fig. 5(f), the detection threshold η_{thresh} is represented by a black horizontal line, while $T^s(\mathbf{X})$ values below the noise level η_{noise} are illustrated with black dots. Events where $T^s(\mathbf{X}) > \eta_{\text{thresh}}$ are highlighted in red, while green represents events that fail to meet the event duration requirement in $T^{s*}(\mathbf{X})$. The evolution from Figs. 5(b) to 5(f) shows significant improvement in humpback unit detectability: Choosing a threshold value that would include all six humpback units in Fig. 5(b) would include a significant amount of shipping noise, while a threshold in Fig. 5(f) can be chosen to include all six humpback units with no inclusion of shipping noise.

The start time, end time, and duration for all events that meet detection requirements are recorded in a log file. A human analyst then prunes false detections from the log file. To aid operator review of the detections in an efficient manner, a graphical user interface (GUI) was designed. The GUI provides a tool for the operator to review time-condensed spectrograms containing the detections, to listen to the detections with adjustable band-passed audio, and to accept or reject each detection. The resulting subset of operator-selected detections can later be used for additional classification.

V. MONTE CARLO SIMULATIONS

To quantify the performance of GPL with known signals over a range of SNR, Monte Carlo simulations were conducted, and the GPL algorithm performance was compared with Nuttall's original power-law processor, two types of energy detection methods, Erbe and King's entropy method, and trained human analysts.

Simulations were considered for three types of noise environments: wind dominated (case 1), distant shipping (case 2), and local shipping (case 3). Case 1 approximates the circumstance of $H_0 = n(t)$, while cases 2 and 3 reflect $H_0 = n(t) + s_1(t)$ with variation in relative contribution of single ship noise, $s_1(t)$, to the total noise field. It is worth noting that case 3 is composed of shipping events recorded in the Santa Barbara channel when one or more large freight vessels were within 5 km of the HARP recording package (depth = 580 m). Six humpback units were selected that spanned varying frequency and temporal ranges in an attempt to characterize detector performance for the wide variety of humpback call units typically seen in acoustic recordings. Ninety-minute segments for each type of noise environment were selected from HARP data free of detectable humpback vocalizations and HARP self-noise. The six characteristic call

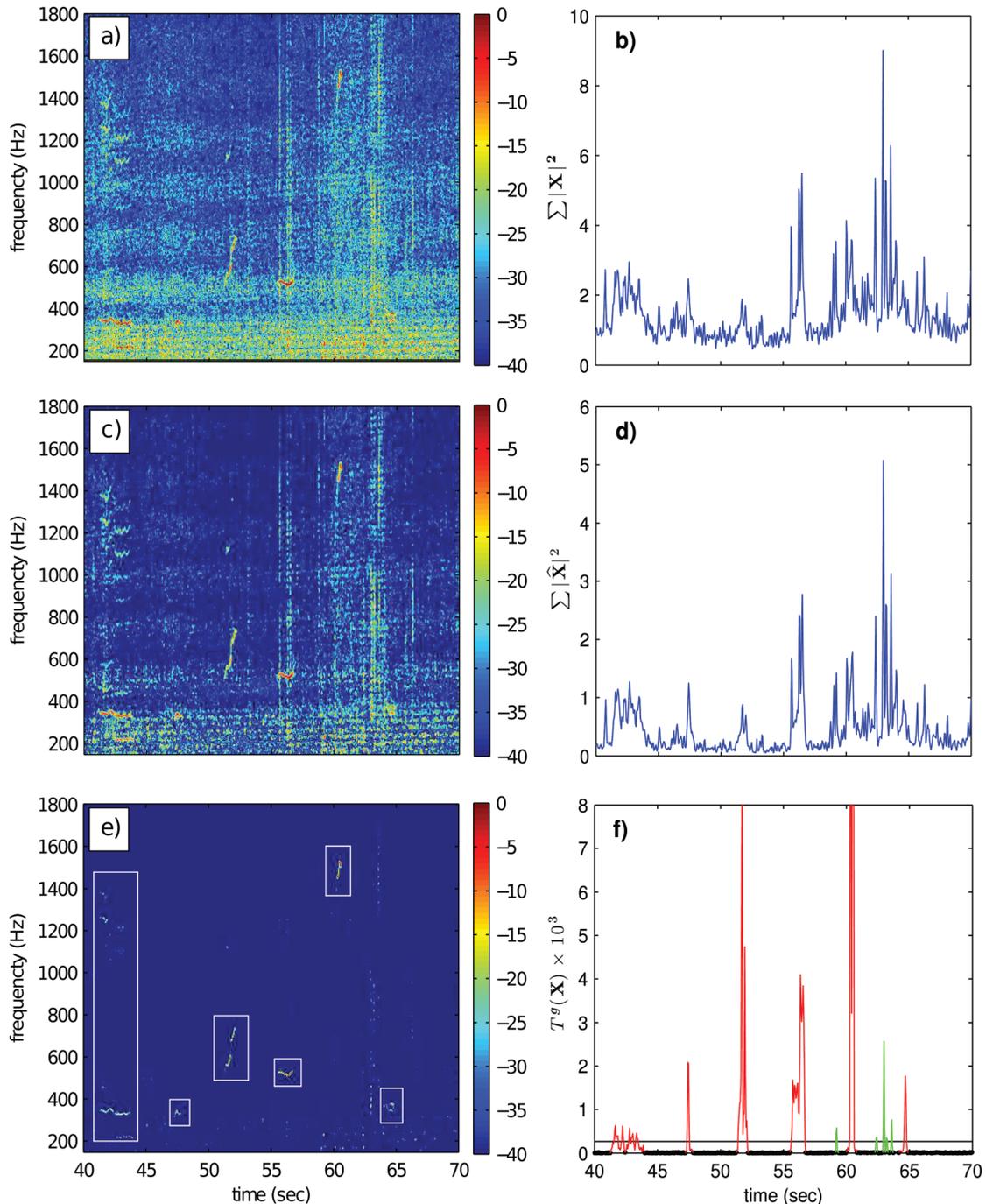


FIG. 5. Visual comparison of energy and GPL for six humpback call units in the presence of local shipping noise starting with (a) conventional spectrogram ($|\mathbf{X}|$) and (b) resulting energy sum, (c) energy with whitener ($|\hat{\mathbf{X}}|$), (d) resulting sum, and finally (e) \mathbf{N} as defined in Sec. III and (f) GPL detector output $T^g(\mathbf{X})$. Units are highlighted in (e) with white boxes. GPL detector output in (f) shows eight groupings of detector statistic values above threshold (horizontal line). The six whale call units (red) meet the minimum time requirements, but the four detections (green) resulting from shipping noise do not, and so are not considered detections. All grams in units of normalized magnitude (dB).

units (shown in Fig. 6) were selected from a different HARP data set that contained humpback vocalizations with high SNR. Noise in these recordings was further reduced using a masking filter in the Fourier domain and then converted back to the time domain to ensure that broadband background noise was not included in the signals of interest. Scaloping (spectral modulation) was avoided by using windows with 93.75% overlap, dividing out the window amplitude in each filtered STFT segment and overlapping successive central segments

by 50%.²² Call units were added in the time domain to a random section of noise for each noise condition. Detection results were recorded for each detection method as described in Kay,²³ using the binary hypothesis test in Eq. (3). Following Kay's example, the observation interval is defined as the duration of the humpback unit of interest. When appropriate, DET curves were created to compare the performance of each detector with varying SNR, where SNR is defined as

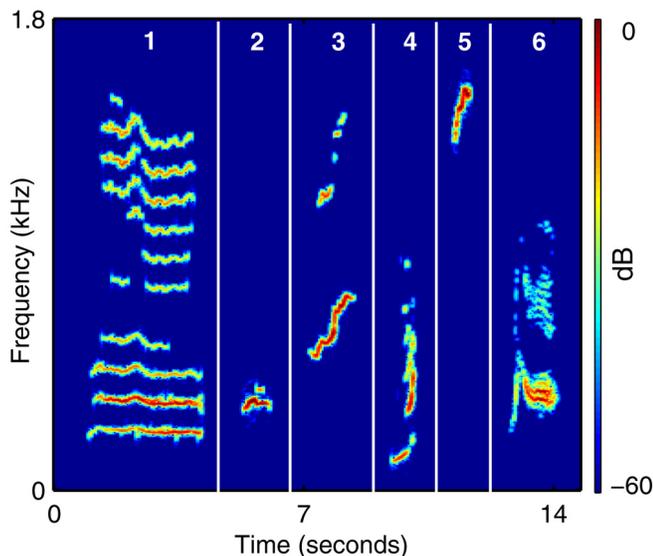


FIG. 6. (Color online) Six humpback units used in Monte Carlo simulations.

$$SNR = 10 \log_{10} \frac{\langle p_s^2 \rangle}{\langle p_n^2 \rangle}$$

where

$$\langle p_s^2 \rangle \equiv \frac{1}{T} \int_0^T p_s^2(t) dt$$

and where p represents the recorded pressure of the time series, band-pass filtered between 150 and 1800 Hz, and T is the duration of the signal. Note that negative SNR in the time domain does not imply negative SNR for individual frequencies following a transformation into the Fourier domain. Detection error tradeoff curves are plots of the two error types from the binary hypothesis test: missed detections (P_{MD}) versus false alarms (P_{FA}). These error types are plotted as a function of detection threshold. DET curves are preferred over traditional receiver operator characteristic (ROC) curves²³ because the missed detection and false alarm axes are scaled to normal distribution fits of the scores of segments with and without signal. DET curves make use of the entire plotting space and are more capable of showing detail when comparing well-performing systems. Best detector performance in the DET space is represented by the point in the lower left corner of DET plots, where the P_{MD} is 0.05% and the P_{FA} is also 0.05%. The point in upper right corner of the plot represents no skill in the detector.

A. Simulations comparing detector performance

In addition to the entropy method described by Erbe and King, two types of energy detectors were included in the analysis. Detector $E^{(1)}$ is defined as a simple energy sum over the frequency range of 150–1800 Hz, which is the equivalent to Nuttall's power-law processor described in Eq. (2) with $\nu = 1$. Assuming an approximate duration of the signal is known, $E^{(1)}$ can be enhanced by using a split window approach.²⁵ Detector $E^{(2)}$ represents this modified approach, as indicated in Eq. (34). For most units, $E^{(2)}$

performs optimally when the number of signal snapshots m_0 corresponds to one-third the signal duration and the number of background snapshots M spans 20 s.

$$E_j^{(2)} = \frac{\sum_{m=-m_0}^{m_0} E_{j+m}^{(1)}}{\sum_{m=-M}^M E_{j+m}^{(1)} - \sum_{m=-m_0}^{m_0} E_{j+m}^{(1)}}. \quad (34)$$

The value of m_0 was adjusted for each unit type during the Monte Carlo simulations, but in practice a single m_0 value would likely be chosen. Additionally, closely spaced call units were not in the simulations, allowing $E^{(2)}$ to perform at its best. Nuttall's power-law processor $T(\mathbf{X})$ was included in the analysis with an exponent $\nu = 3$, which was found to be the optimal exponent for the simulations. Simulations for GPL were conducted with and without the parameter metric enhancements $T^{g*}(\mathbf{X})$.

To minimize the influence of the whitener, both energy methods and the entropy method used the conditional whitener prescribed in Eqs. (10) and (11), as it increased performance for all three methods. The conditional whitener was not used with Nuttall's original power-law processor as it decreased performance.

For each of the detectors, Monte Carlo simulations were conducted for all six unit types in Fig. 6, with SNR ranging from -10 to 10 dB, and noise cases 1-3. Based on examination of trained human analysts' picks, a SNR of -3 dB corresponds to a human P_{MD} of approximately 15% in case 1, 18% in case 2, and over 20% for case 3. The detector DET statistics for units 1–6 were combined and are shown for each detector in Fig. 7 with 10,000 trials for each unit, noise condition and SNR. The GPL test statistic $T^g(\mathbf{X})$ is shown in preference to $T^{g*}(\mathbf{X})$ to put all the detection algorithms on an equal footing. In noise case 1, all detection methods meet the inequality constraints in Eq. (5). In noise case 2, both $T(\mathbf{X})$ and $T^g(\mathbf{X})$ meet the constraints. In noise case 3, only $T^g(\mathbf{X})$ satisfies the constraints. The DET statistics do not address the stability of η_{thresh} among noise conditions, which is discussed further in succeeding sections. It is worth noting that the performance of $E^{(2)}$ is susceptible to considerable performance degradation when the short-term averaging duration is not selected carefully. In wind-driven noise conditions, it is found that a simple energy sum often has better detector performance than $E^{(2)}$. However, in the presence of shipping noise, detection method $E^{(2)}$ consistently outperformed $E^{(1)}$.

Table II summarizes the GPL threshold DET statistics using the parameter enhancement $T^{g*}(\mathbf{X})$ for all call units and noise conditions over a range of SNR using the defined value for η_{thresh} . Threshold DET statistics are not provided for the other detection techniques because they do not satisfy the inequality constraints, and also establishing appropriate threshold values is somewhat arbitrary. GPL had nearly perfect detection scores for all six unit types in all three noise cases for SNR of 0 dB and higher. For SNR -2 dB, GPL had P_{MD} below 2% for all unit types and noise cases except unit 4. The majority of energy in unit 4 is contained within a very

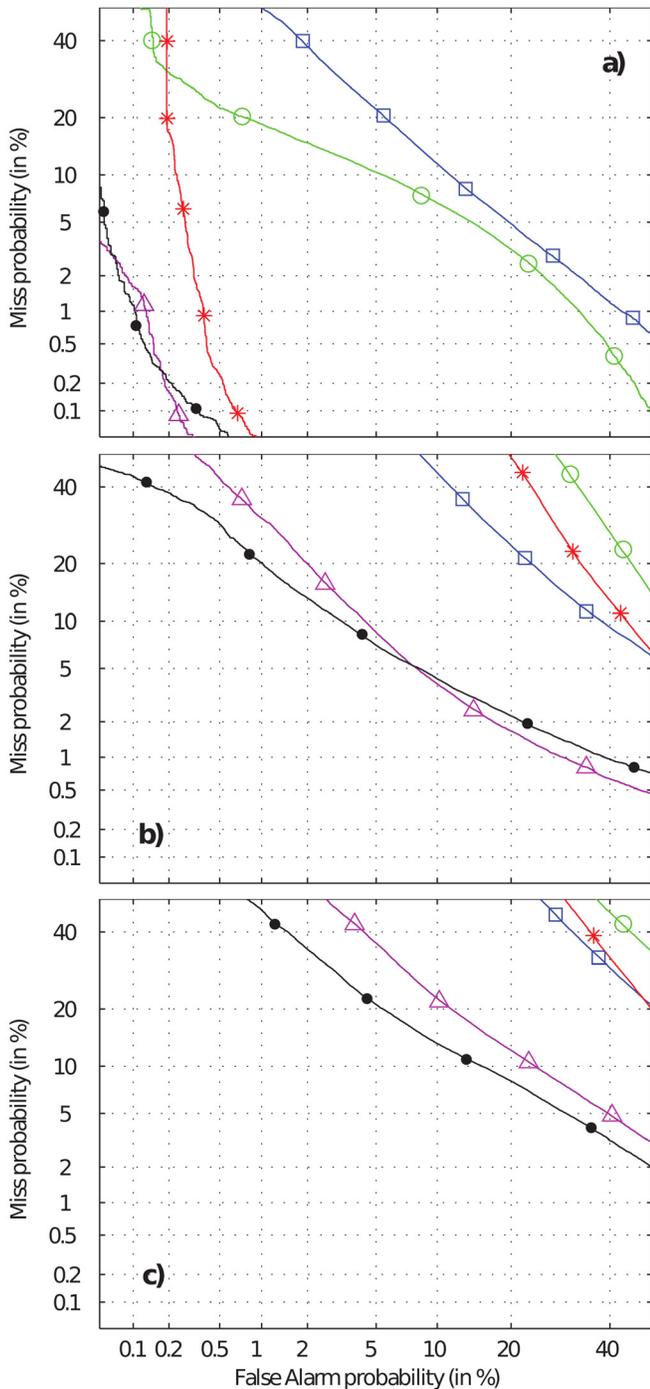


FIG. 7. (Color online) DET results for units 1–6 with SNR -3 dB in noise dominated by (a) wind-driven noise, (b) distant shipping, and (c) local shipping, for GPL (closed circle), Nuttall (open triangle), entropy (asterisk), $E^{(1)}$ (open circle), and $E^{(2)}$ (open square).

narrow time interval of 0.3 s. Therefore unit 4 required slightly higher SNR than the rest of the unit types to consistently meet the minimum event duration requirement. It is also worth noting that the DET statistics are better in cases 2 and 3 than case 1 in very low SNR conditions. Because SNR is defined as the ratio of time-integrated squared pressure band-limited between 150 and 1800 Hz, the low frequency distribution of noise in cases 2 and 3 can allow for locally higher SNR in the frequency bands in which the unit occurs

and results in an increase in detectability for very low SNR units. In general, units with the shortest durations, lowest frequencies, and units lacking frequency sweeps prove hardest to detect using the GPL algorithm. This result is expected because units at low SNR with very short duration may be rejected for failing to meet τ_c . Low frequency units tend to be more susceptible to masking by shipping, and monotone units are more liable to be suppressed during normalization. The first two weaknesses in detection are also shared by human analysts, the third applies to GPL alone.

Humpback call analysts would like the ability to categorize humpback song into types of units. To this end, Table II will help provide guidelines for minimum SNR conditions that should be met before the detector can reliably detect all humpback units. The augmented model parameters $[\Theta, \eta_{\text{thresh}}, \eta_{\text{noise}}, \tau_c]$ were found to be robust for 2 yr of data analyzed at multiple locations throughout the southern California Bight, the coast of Washington state, and Hawaii. However, these values may need to be adjusted slightly if ocean noise conditions change appreciably from the noise recorded at these locations. Hydrophones located at shallower depths, sea ice noise, and the presence of noise generated from oil exploration are some circumstances that may warrant adjustments.

B. Simulations comparing power-law detectors to trained human analysts

A second set of simulations was conducted to compare the performance of $T^{*}(\mathbf{X})$ and Nuttall's test statistic $T(\mathbf{X})$ with trained human analysts. Here, five additional humpback units were included with the original six units shown in Fig. 6 to prevent the operators from recognizing repeated units. These 11 units were inserted into the 90-min recordings of cases 1–3 with varying SNR, totaling 220 units for each of the three noise conditions. Each human analyst was asked to identify all humpback units and was not told the number, locations, or SNR of the signals present. The GPL P_{MD} values were calculated using the standard value of η_{thresh} , which was chosen so that $P_{FA} < P_{FA}^{max}$ for the strongest shipping conditions. The results using this threshold, shown in Table III, illustrate that the GPL algorithm was

TABLE II. Probability of missed detection and probability of false alarm (P_{MD}/P_{FA} , given as percentage) using η_{thresh} for units 1–6, varying SNR and noise cases, 10,000 trials per statistic.

SNR	Noise	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5	Unit 6
-6 dB	Case 1	98.5/1.0	87.2/0.0	98.2/0.0	100/0.0	98.9/0.0	95.4/0.0
	Case 2	87.9/4.8	77.7/4.7	84.0/4.9	94.7/4.5	78.8/4.1	89.6/4.5
	Case 3	78.5/6.0	81.6/5.7	73.1/6.5	92.1/5.7	31.6/5.0	83.2/4.7
-4 dB	Case 1	18.7/0.0	14.8/0.0	8.0/0.0	98.8/0.0	10.2/0.0	0.7/0.0
	Case 2	21.5/5.2	10.6/4.5	1.9/4.7	92.7/3.8	0.4/4.2	16.7/4.6
	Case 3	32.3/6.3	26.2/5.7	4.0/6.1	89.3/5.3	0.0/4.8	39.3/6.8
-2 dB	Case 1	0.0/0.0	0.0/0.0	0.0/0.0	23.8/0.0	0.0/0.0	0.0/0.0
	Case 2	0.1/5.0	0.1/4.3	0.0/4.9	47.0/4.1	0.0/4.2	0.2/4.8
	Case 3	0.0/6.9	0.6/5.6	0.0/6.6	62.2/5.3	0.0/5.2	1.6/6.5
0 dB	Case 1	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0	0.0/0.0
	Case 2	0.0/5.1	0.0/4.4	0.0/4.8	3.4/4.4	0.0/4.5	0.0/5.1
	Case 3	0.0/6.3	0.0/5.3	0.0/6.7	0.0/5.5	0.0/5.0	0.0/6.4

able to detect lower SNR signals slightly better than the human analysts and performed roughly on a par with the human analysts for higher SNR. Each operator was able to improve their performance by reviewing the output of the GPL detector.

For comparison purposes, Eq. (2) with $\nu=3$ was included in Table III to show the performance of a constant threshold using Nuttall's original power-law processor. A threshold was chosen using the same construction as for GPL, shown in Fig. 3, limiting the relative proportion of false detections in case 3 to the same level. In doing so, the P_{MD} for cases 1 and 2 violate the constraints stated in Eqs. (4) and (5) as humans were able to identify a significantly higher number of units at low SNR. For this reason, Eq. (2) is not further considered.

VI. PARAMETER ESTIMATION

In addition to detecting the presence or absence of a humpback unit, it is often desired to mark the beginning and end times of the humpback unit in the time series. If this can be done automatically and accurately, then that unit can be selected from the time series and passed to a classification scheme that can measure additional metrics about the unit. Even without further classification, unit timing parameters are provided by GPL itself, providing useful statistics on call rate, repetition, and both short-term and long-term calling trends. Parameter estimation algorithms and human analysts may provide different start and end time estimates for the same call unit depending on the noise condition and SNR. As SNR decreases, the edges of the unit may often be indistinguishable from the noise, and so a human analyst or automated algorithm tends to mark a shorter unit duration at lower SNR, even when the vocalizing source is producing a unit with the same duration in both cases. Additionally, all three detectors and human analysts are subject to the limitations imposed by the STFT length and window overlap as previously discussed. The bias and standard deviation in estimating unit duration are documented in this section for the GPL algorithm over a range of SNR, noise conditions, and unit types. Using the same six unit types from the Monte

TABLE III. Probability of missed detection (P_{MD} , given as a percentage) for GPL versus baseline power-law detector (Nuttall) and human analysts for varying SNR. Detector threshold values were established such that Case 3 $P_{FA} < 6\%$ and applied to cases 1 and 2.

SNR		-6 dB	-4 dB	-2 dB	0 dB
Case 1	GPL	74.6	10.9	10.9	0.0
	Nuttall	94.6	32.7	10.9	0.0
	Analyst 1	74.6	21.8	12.7	3.6
	Analyst 2	76.4	18.2	9.1	5.4
Case 2	GPL	60.0	14.6	12.7	7.3
	Nuttall	81.8	41.8	14.6	7.3
	Analyst 1	78.0	24.0	12.0	6.0
	Analyst 2	81.9	27.3	10.9	7.3
Case 3	GPL	61.8	27.3	9.1	5.5
	Nuttall	61.8	29.1	7.3	3.6
	Analyst 1	84.0	48.0	14.0	14.0
	Analyst 2	56.4	23.7	7.3	3.7

Carlo simulations, the units were inserted into the three noise conditions with SNR varying from -4 to 10 dB, with 500 trials per condition. For comparison, the two energy detectors were also included in this analysis, where the unit duration was marked by the time that passed in which the energy of the unit was above threshold. This method is similar to that used in ISHMAEL,⁹ in which the user is able to extract time series segments for calls that pass the user-defined threshold. For consistency in comparison with GPL, a threshold value for the energy techniques was chosen in which on average the P_{MD} was 10% for call units 1-6 for noise case 1, with SNR of -2 dB. For noise case 1, an SNR of -2 dB was sufficiently high for a human to consistently and accurately detect nearly all call units in the record. The threshold and baseline values for marking call units with the GPL algorithm remained consistent with those described in Sec. IV.

Table IV shows call duration parameters for units 1 and 3 with unit 1 representing the most error in parameter estimation for GPL, while unit 3 represents typical performance. The quantity Δt_s represents the bias of the estimated unit start time in seconds from the true unit start time ($\hat{t}_s - t_s^{true}$), σ_s represents the standard deviation of \hat{t}_s . Likewise, the quantity Δt_e represents the bias in seconds of the unit end time estimate ($\hat{t}_e - t_e^{true}$), and σ_e represents the standard deviation for \hat{t}_e .

For units greater than 2 dB SNR in noise cases 1 and 2, GPL is able to accurately measure start and end times, with Δt_s and Δt_e at 0.09 s or smaller and both σ_s and σ_e at 0.10 s or smaller. The two energy methods are also fairly effective at measuring these parameters at 2 dB or higher in noise case 1. $E^{(1)}$ is not useful in either noise case 2 or 3 because the threshold chosen for $E^{(1)}$ to work well in noise case 1 creates large overestimates when ship noise is present. While at first glance $E^{(2)}$ appears to also work well in noise cases 2 and 3, using the threshold optimized for noise case 1 results in many false alarms. Raising the threshold reduces P_{FA} , but unit durations are then drastically underestimated and the standard deviation is large.

VII. OBSERVATIONAL RESULTS

The performance of GPL using $T^s*(\mathbf{X})$ was established for three HARP deployments with varying humpback unit structure, SNR, depth, and noise conditions. Although the entropy detector, Nuttall's original power-law processor, and the energy methods violate the constraints in Eq. (5), $E^{(1)}$ and $E^{(2)}$ were included in the observational results because of their prevalence in marine mammal detection software. Twenty hours of acoustic recordings were first examined by trained human analysts, and humpback call units were identified for each of three locations off the California coast. Additionally, operators reviewed the detections produced by GPL and energy-based methods to include any units first missed by the operators but captured by the detectors. Unlike the Monte Carlo simulations where the humpback unit locations are known regardless of signal strength, in the observational data, the locations of humpback units are only known within the detection ability of a trained operator. This operator-derived information was used as ground truth. As in the Monte Carlo simulations, binary hypothesis test metrics are

TABLE IV. Start-time bias Δt_s , end time bias Δt_e , start time standard deviation σ_s , and end time stand deviation σ_e in seconds for unit 1 (duration; 3.34 s) and unit 3 (duration; 1.3 s).

Unit 1	Type	Noise case 1				Noise case 2				Noise case 3			
		Δt_s	σ_s	Δt_e	σ_e	Δt_s	σ_s	Δt_e	σ_e	Δt_s	σ_s	Δt_e	σ_e
-2 dB	E1	-1.38	0.63	-0.62	0.50	-0.78	2.27	-0.66	3.65	22.22	21.41	23.83	22.33
	E2	-1.00	0.41	-0.71	0.27	-0.96	0.55	-0.84	0.54	-1.00	0.71	-0.85	0.69
	GPL	-0.34	0.08	-0.02	0.06	-0.35	0.17	-0.16	0.33	-0.34	0.20	-0.19	0.28
0 dB	E1	-0.49	0.21	-0.23	0.10	-0.48	3.48	0.14	3.29	22.71	21.92	23.43	22.26
	E2	-0.43	0.06	-0.39	0.06	-0.46	0.22	-0.43	0.23	-0.50	0.35	-0.44	0.32
	GPL	-0.21	0.10	0.01	0.03	-0.21	0.14	-0.02	0.11	-0.22	0.14	-0.02	0.11
2 dB	E1	-0.31	0.10	-0.15	0.03	0.29	3.54	0.63	3.84	20.63	20.64	25.36	23.06
	E2	-0.28	0.04	-0.23	0.04	-0.29	0.10	-0.25	0.09	-0.29	0.15	-0.25	0.15
	GPL	-0.09	0.05	0.03	0.03	-0.09	0.10	0.02	0.08	-0.09	0.09	0.03	0.10
Unit 3	Type	Δt_s	σ_s	Δt_e	σ_e	Δt_s	σ_s	Δt_e	σ_e	Δt_s	σ_s	Δt_e	σ_e
-2 dB	E1	-0.46	0.21	-0.36	0.16	0.26	3.63	0.34	4.28	23.04	22.36	23.95	23.18
	E2	-0.39	0.15	-0.47	0.19	-0.36	0.22	-0.41	0.18	-0.33	0.32	-0.36	0.33
	GPL	-0.01	0.05	0.01	0.04	0.02	0.16	0.04	0.15	0.00	0.12	0.05	0.13
0 dB	E1	-0.20	0.09	-0.20	0.04	0.43	4.31	0.59	4.49	22.46	22.62	22.58	22.58
	E2	-0.22	0.09	-0.29	0.06	-0.21	0.19	-0.29	0.14	-0.21	0.24	-0.29	0.23
	GPL	0.03	0.04	0.05	0.04	0.06	0.31	0.09	0.41	0.06	0.11	0.07	0.12
2 dB	E1	-0.11	0.03	-0.15	0.03	0.52	3.64	0.28	2.51	24.15	22.25	23.70	22.14
	E2	-0.07	0.05	-0.21	0.03	-0.08	0.10	-0.21	0.06	-0.09	0.18	-0.20	0.18
	GPL	0.06	0.04	0.07	0.03	0.07	0.08	0.08	0.06	0.08	0.11	0.10	0.12

used to evaluate the detector performances. An observation interval of 3 s is used for determining the detector output. Specifically, the maximum value of each detector output is recorded in a 3 s window surrounding each known humpback unit. The portions of the acoustic record that contained only noise are also broken into 3 s observation windows. The maximum detector output is recorded for each noise observation window using the same method as the signal-present windows. DET curves were produced for each of the three HARP deployments for GPL, $E^{(1)}$, and $E^{(2)}$.

Site SurRidge is 50 km southwest of Monterey, and the recording package is at a depth of 1386 m. Site B, located inside the Santa Barbara shipping channel, is 25 km north of Santa Rosa Island and the recording package is at a depth of 580 m. Site N is located 50 km southwest of San Clemente Island and contains a recording package at a depth of 750 m.

Figure 8(a) shows the DET curves for 20 h of duty cycled acoustic recordings at site SurRidge spanning January 26–28, 2008. The analysis period contains 1041 humpback call units with most units categorized as low SNR with few identifiable harmonics. Local shipping noise is dominant during 14% of the record, distant shipping is dominant during 62% of the record, and wind-driven noise is dominant during 24% of the record. Both $E^{(1)}$ and $E^{(2)}$ perform poorly during this period, with $E^{(1)}$ performing worse than $E^{(2)}$. The GPL algorithm performs reasonably well and is able to detect all the units marked by the operator with a 4% P_{FA} .

Figure 8(b) shows the DET curves for 20 h of duty cycled recordings at site B spanning April 16–18, 2008. The analysis period contains 4546 humpback call units with most units categorized as moderate SNR with occasional calling bouts with high SNR. Local shipping noise is dominant during 36% of the record, distant shipping is dominant

during 59% of the record, and wind-driven noise is dominant during 5% of the record. Both $E^{(1)}$ and $E^{(2)}$ perform poorly during this period with $E^{(1)}$ performing worse than $E^{(2)}$. The GPL algorithm performs well and is able to detect all the units marked by the operator with just over 2% P_{FA} .

Figure 8(c) shows the DET curves for 20 h of continuous recordings at site N spanning December 6–7, 2009. The analysis period contains 15 450 humpback call units with most units categorized as high SNR containing many harmonics with occasional calling at low SNR. Local shipping noise is dominant during 15% of the record, distant shipping is dominant during 23% of the record, and wind-driven noise is dominant during 62% of the record. The detector $E^{(1)}$ performs better than $E^{(2)}$ in this scenario, which can be attributed to the extremely high call rate for this recording. Because $E^{(2)}$ uses a short-term average compared with a long-term average, units in close proximity often decrease the detector output. Because the GPL algorithm uses an iterative strategy in determining units, it is less affected by high calling rates. Therefore the GPL algorithm outperforms $E^{(1)}$ and $E^{(2)}$ by a wide margin in this environment, detecting every unit marked by the operator with just over 0.5% P_{FA} .

Each deployment contains a handful of questionable humpback signals. When the questionable signals are included as units, the P_{MD} becomes nonzero but remains 2% or less for each deployment.

At first glance, the steep vertical slope of the DET curve for GPL performance in Fig. 8 can lead to the conclusion of an unstable detection threshold because a seemingly small change in P_{FA} appears to have a large effect on P_{MD} . The reason for this steep slope is twofold: Using the statistic $T^{g*}(\mathbf{X})$ instead of $T^g(\mathbf{X})$ enhances the non-Gaussian distribution of the test statistic, as shown in the histogram in Fig. 9. Here one can see that a

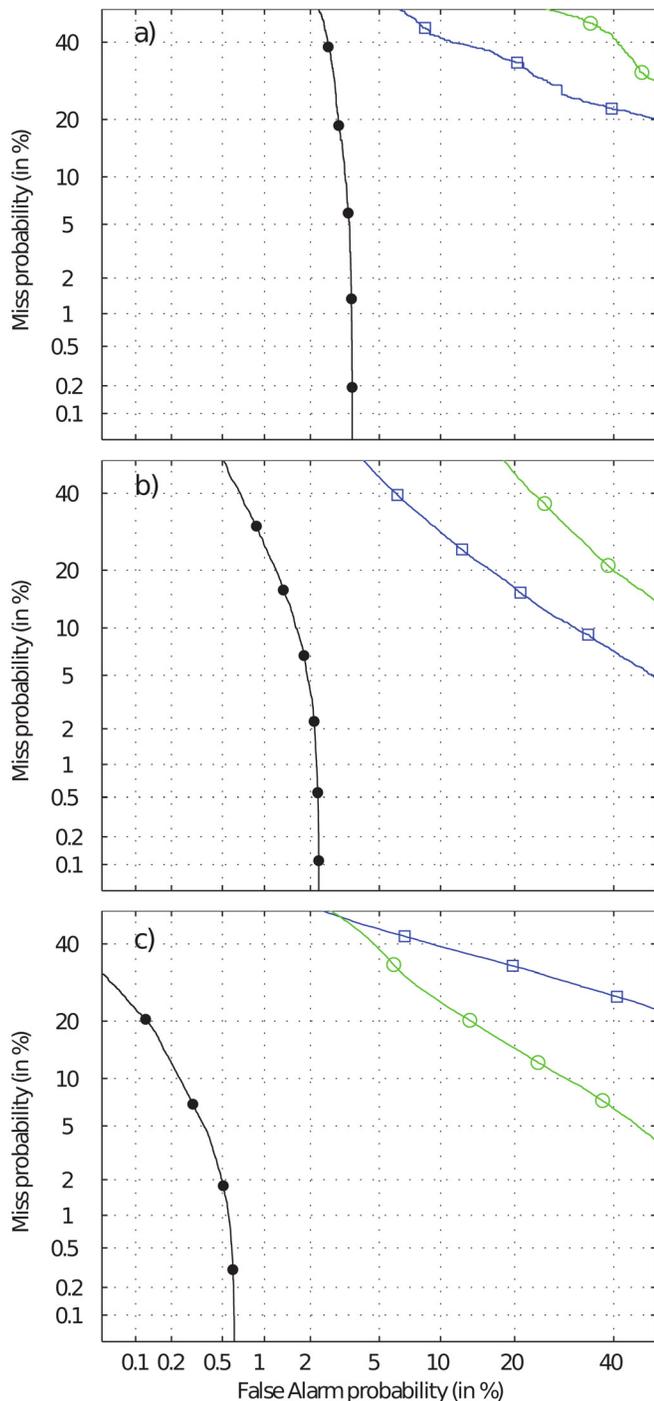


FIG. 8. (Color online) DET results for HARP deployments at (a) Site SurRidge, (b) Site B, and (c) Site N for GPL (closed circle), energy sums $E^{(1)}$ (open circle), and $E^{(2)}$ (open square).

vast majority of events have detector output values of zero because detections that do not meet the τ_c duration requirement are forced to zero. This binary decision within the GPL logic creates a sharp but stable elbow in the DET curve. Additionally, low SNR units that would have received low values of $T^{g^*}(\mathbf{X})$ were not identified by human analysts, which also alters the shape of the DET curves as compared to Fig. 7.

To evaluate the stability in the GPL threshold value among the three HARP deployments, the P_{FA} and P_{MD} are calculated using the standard threshold of $\eta_{\text{thresh}} = 2.62 \times 10^{-4}$.

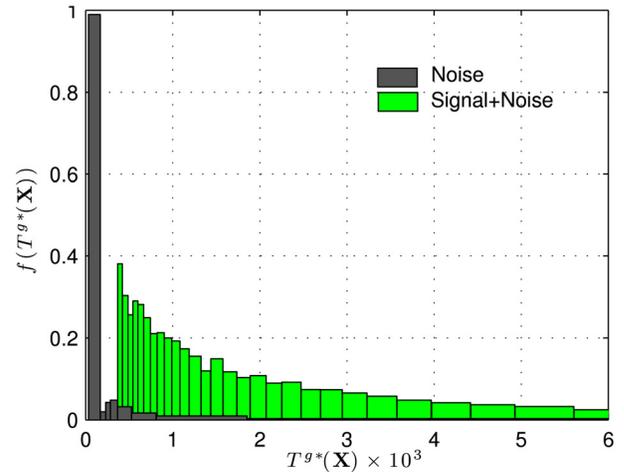


FIG. 9. (Color online) Normalized histogram of detector outputs for signal and signal + noise for Site N deployment.

Site SurRidge had $P_{FA} = 3.7\%$ and $P_{MD} = 0\%$, site N had $P_{FA} = 1.1\%$ and $P_{MD} = 0\%$, and site B had $P_{FA} = 3.2\%$ and $P_{MD} = 0\%$. These results suggest that the chosen value of η_{thresh} is both a stable and a sensible choice for all three HARP deployments despite varying signal and noise conditions. Undoubtedly, the GPL algorithm misses some humpback units that occurred in these records. However, because human analysts are used to establish a ground truth of humpback unit occurrences, the low P_{MD} values verify that the GPL algorithm is able to find nearly all units that could be verified by human analysts.

VIII. CONCLUSIONS

The generalized power-law processor outperforms energy detection techniques for finding humpback vocalizations in the presence of shipping noise and wind-generated noise in the southern California Bight. The normalization over both frequency and time permits fixed thresholds that can be used throughout long deployments having varying ocean noise conditions. The algorithm capitalizes on basic parameters of the signal and noise environments yet remains general enough to capture all types of humpback units without the need for predefined templates. The detector is designed to capture all humpback units that are detectable by trained human analysts, while maintaining a low probability of false alarms. The detector performance was verified by inserting humpback units with varying SNR into three noise conditions and comparing the detector output to that of two trained operators. Additionally, the GPL algorithm is able to detect nearly all humpback units previously identified by human analysts in three different deployments off the coast of California with a result of $P_{FA} = 3.7\%$ or better. This performance allows a human analyst to review a much smaller subset of data when looking for humpback units.

Once the periods of data containing humpback units have been identified, basic call parameters such as unit duration, center frequency, number of units, and inter-call interval can be automatically tabulated. The GPL process provides considerably more detail than basic presence/absence metrics to

which human analysis is typically restricted, owing to the labor intensive nature of manually selecting individual units. Parameter estimation performance obtained from simulations show that GPL commonly yields precision of 0.1 s or less for estimating the beginning and end of a unit for reasonable SNR under all but heavy shipping noise. By contrast, measuring unit duration parameters using energy detection techniques proved unfeasible except in high SNR situations. Although the analysis here has focused on algorithm settings tuned to the specific characteristics of humpback vocalizations, the GPL algorithm has in fact the potential to be modified for many types of marine mammal vocalizations and is likely to prove useful as a precursor to classification techniques.

ACKNOWLEDGMENTS

The authors are extremely grateful to Greg Campbell, Amanda Cummins, and Sara Kerosky, who provided operator-identified humpback whale unit locations and trained human analyst expertise. Special thanks to Sean Wiggins and the entire Scripps Whale Acoustics lab for providing thousands of hours of high quality acoustic recordings. Bill Hodgkiss was extremely helpful in providing feedback in areas of signal processing, Monte Carlo simulations, and detection theory. The authors are grateful to Peter Rickwood, who at the early stages in this work provided time, expertise, and software in our initial evaluation of schemes for classification. The first author would like to thank the Department of Defense Science, Mathematics and Research for Transformation Scholarship program, the Space and Naval Warfare Systems Command Center (SPAWAR) Pacific In-House Laboratory Independent Research program, and Rich Arrieta from the SPAWAR Unmanned Maritime Vehicles Lab for continued financial and technical support. Work was also supported by the Office of Naval Research, Code 32, CNO N45, and the Naval Postgraduate School.

APPENDIX: MATHEMATICAL DETAILS

The numerator in Eq. (14) has a pdf of $\chi_{K-1}^2(z)$ and the denominator $\chi_2^2(z)$ so the quantity $X/(K-1)$ is thus an F -distribution of the form

$$f_X(x) = \left(\frac{(K-1)x}{1+(K-1)x} \right)^{K-2} \left(\frac{K-1}{1+(K-1)x} \right)^2. \quad (\text{A1})$$

Observe that

$$\begin{aligned} P(Y < y) &= P(X > (K-1)^{-1}(1/y-1)) \\ &= 1 - F_X((K-1)^{-1}(1/y-1)), \end{aligned}$$

accordingly

$$\begin{aligned} f_Y(y) &= \frac{1}{y^2} f_X((K-1)^{-1}(1/y-1)) \\ &= (K-1)(1-y)^{K-2} \end{aligned} \quad (\text{A2})$$

and therefore

$$F_Y(y) = 1 - (1-y)^{K-1}.$$

With the statistics of entries in \mathbf{A} thus characterized, it is logical to try to extend this line of reasoning to the product form of Eq. (6) by attempting first to reproduce the equivalent of Eq. (15). For simplicity, consider $J=K$ and $\gamma=1$. Then the reciprocal leads to a homogeneous form $1+Z_1+Z_2$ where

$$\begin{aligned} Z_1 &= \frac{\sum_{n=1}^{K'} |X_{n,j}|^2 + \sum_{m=1}^{K'} |X_{k,m}|^2}{|X_{k,j}|^2}, \\ Z_2 &= \frac{\sum_{n=1}^{K'} |X_{n,j}|^2 \sum_{m=1}^{K'} |X_{k,m}|^2}{|X_{k,j}|^4}. \end{aligned} \quad (\text{A3})$$

The first term in Eq. (A4) is another F -distribution as in Eq. (A1) but with K replaced by $2K$. The difficulty comes from the second term. For the second term, the pdfs for its numerator and denominator are

$$\frac{(2K-3)z^{K-2}}{\Gamma(K-1/2)^2} \mathcal{K}_1(2\sqrt{z}) \quad \text{and} \quad \frac{1}{2} z^{-1/2} e^{-z^{1/2}},$$

respectively, where \mathcal{K} is the modified Bessel function of the second kind. This ratio is not an F -distribution and appears not to be characterized. Thus even for this first extension of normalization beyond Eq. (13), immediate recourse to asymptotic approximation is necessary.

Last, for the pdf governing Eq. (19), it is immediate on a change of variable that

$$\begin{aligned} f_Z^{(p)}(z) &= \frac{2}{pz^{(p-1)/p}} (\sqrt{\pi}/2 + \sqrt{z}) e^{-(\sqrt{\pi}/2 + \sqrt{z})^2}, \\ z &> \pi^{p/2}/2^p, \end{aligned} \quad (\text{A4})$$

and the symmetric combination $f_Z^{(p)}(z) + f_Z^{(p)}(-z)$ applies for $0 \leq z \leq \pi^{p/2}/2^p$ to account for both roots in that interval.

¹R. Payne and S. McVay, "Songs of humpback whales," *Science* **173**, 585–597 (1971).

²S. Cerchio, J. Jacobsen, and T. Norris, "Temporal and geographical variation in songs of humpback whales, *Megaptera novaeangliae*: Synchronous change in Hawaiian and Mexican breeding assemblages," *Anim. Behav.* **62**, 313–329 (2001).

³D. Mellinger and C. Clark, "Recognizing transient low-frequency whale sounds by spectrogram correlation," *J. Acoust. Soc. Am.* **107**, 3518–3529 (2000).

⁴J. Potter, D. Mellinger, and C. Clark, "Marine mammal call discrimination using artificial neural networks," *J. Acoust. Soc. Am.* **96**, 1255–1262 (1994).

⁵J. Brown and P. Smaragdis, "Hidden Markov and Gaussian mixture models for automatic call classification," *J. Acoust. Soc. Am.* **125**, EL221–EL224 (2009).

⁶P. Rickwood and A. Taylor, "Methods for automatically analyzing humpback song units," *J. Acoust. Soc. Am.* **123**, 1763–1772 (2008).

⁷X. Mouy, M. Bahoura, and Y. Simard, "Automatic recognition of fin and blue whale calls for real-time monitoring in the St. Lawrence," *J. Acoust. Soc. Am.* **126**, 2918–2928 (2009).

⁸T. Abbot, V. Premus, and P. Abbot, "A real-time method for autonomous passive acoustic detection-classification of humpback whales," *J. Acoust. Soc. Am.* **127**, 2894–2903 (2010).

- ⁹D. Mellinger, "Ishmael 1.0 users guide," NOAA Technical Memorandum OAR PMEL-120, available from NOAA/PMEL, Seattle, WA (2001).
- ¹⁰H. Figueroa, xbat, Version 5 (Cornell University Bioacoustics Research Program, 2007).
- ¹¹D. Gillespie, D. Mellinger, J. Gordon, D. McLaren, P. Redmond, R. McHugh, P. Trinder, X. Deng, and A. Thode, "PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans," *J. Acoust. Soc. Am.* **125**, 2547–2547 (2009).
- ¹²C. Erbe and A. King, "Automatic detection of marine mammals using information entropy," *J. Acoust. Soc. Am.* **124**, 2833–2840 (2008).
- ¹³A. Nuttall, "Detection performance of power-law processors for random signals of unknown location, structure, extent, and strength," NUWC-NPT Technical Report, Newport, RI (1994).
- ¹⁴A. Nuttall, "Near-optimum detection performance of power-law processors for random signals of unknown locations, structure, extent, and arbitrary strengths," NUWC-NPT Technical Report, Newport, RI (1996).
- ¹⁵S. Wiggins, "Autonomous Acoustic Recording Packages (ARPs) for long-term monitoring of whale sounds," *Marine Tech. Soc. J.* **37**, 13–22 (2003).
- ¹⁶S. Wiggins, M. Roch, and J. Hildebrand, "TRITON software package: Analyzing large passive acoustic monitoring data sets using MATLAB," *J. Acoust. Soc. Am.* **128**, 2299–2299 (2010).
- ¹⁷Z. Wang and P. Willett, "All-purpose and plug-in power-law detectors for transient signals," *IEEE Trans. Signal Process.* **49**, 2454–2466 (2001).
- ¹⁸A. Stuart and K. Ord, *Kendall's Advanced Theory of Statistics. Distribution Theory* (Wiley, New York, 2009), Vol. 1, Chaps. 1–5, 8–11.
- ¹⁹W. Struzinski and E. Lowe, "A performance comparison of four noise background normalization schemes proposed for signal detection systems," *J. Acoust. Soc. Am.* **76**, 1738–1742 (1984).
- ²⁰R. Charif, C. Clark, and K. Fristrup (2004), "RAVEN 1.2 users manual, Appendix B: A Biologists Introduction to Spectrum Analysis," Cornell Laboratory of Ornithology, Ithaca, New York.
- ²¹M. Beecher, "Spectrographic analysis of animal vocalizations: Implications of the uncertainty principle," *Bioacoustics* **1**, 187–208 (1988).
- ²²R. Lowdermilk and F. Harris, "Using the FFT as an arbitrary function generator," in *Proceedings of AUTOTESTCON (2005) (IEEE)*, pp. 408–412.
- ²³S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice-Hall, Englewood Cliffs, NJ, 1998), Vol. 2, pp. 61, 74, 269.
- ²⁴A. Martin, G. Doddington, T. Kamm, M. Ordowski, and M. Przybocki, "The DET curve in assessment of detection task performance," in *Proceedings of Eurospeech*, Rhodes, Greece (1997), Vol. 97, pp. 1895–1898.
- ²⁵R. Nielsen, *Sonar Signal Processing* (Artech House, Inc., Norwood, MA, 1991), pp. 145–147.