



This is a repository copy of *Utilising temporal signal features in adverse noise conditions: Detection, estimation, and the reassigned spectrogram*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/95754/>

Version: Accepted Version

Article:

Mill, R.W. and Brown, G.J. (2016) Utilising temporal signal features in adverse noise conditions: Detection, estimation, and the reassigned spectrogram. *Journal of the Acoustical Society of America*, 139. pp. 904-917. ISSN 1520-8524

<https://doi.org/10.1121/1.4941566>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

**Utilising temporal signal features in adverse noise conditions: Detection,
estimation and the reassigned spectrogram**

Robert W. Mill and Guy J. Brown^{a)}

*University of Sheffield,
Department of Computer Science,
Regent Court,
211 Portobello,
Sheffield,
S1 4DP*

(Dated: January 13, 2016)

Abstract

Visual displays in passive sonar based on the Fourier spectrogram are underpinned by detection models that rely on signal and noise power statistics. Time-frequency representations specialised for sparse signals achieve a sharper signal representation, either by reassigning signal energy based on temporal structure or by conveying temporal structure directly. However, temporal representations involve nonlinear transformations that make it difficult to reason about how they respond to additive noise. This article analyses the effect of noise on temporal fine structure measurements such as zero crossings and instantaneous frequency. Detectors that rely on zero crossing intervals, intervals and peak amplitudes, and instantaneous frequency measurements are developed, and evaluated for the detection of a sinusoid in Gaussian noise, using the power detector as a baseline. Detectors that rely on fine structure outperform the power detector under certain circumstances; and detectors that rely on both fine structure *and* power measurements are superior. Reassigned spectrograms assume that the statistics used to reassign energy are reliable, but the derivation of the fine structure detectors indicates the opposite. The article closes by proposing and demonstrating the concept of a *doubly-reassigned spectrogram*, wherein temporal measurements are reassigned according to a statistical model of the noise background.

PACS numbers: 43.60.Hj, 43.60.Bf

^{a)}Text: g.j.brown@sheffield.ac.uk

I. INTRODUCTION

A wide variety of acoustic signals consist of spectrotemporally “sparse” modulated narrowband components buried in broadband noise. Familiar examples include speech in traffic noise, or birdsong in wind blowing through rustling leaves. Many mechanical emissions, such as those of marine vessels, also belong to this category: periodic sources, such as rotating machine parts and alternating electrical currents, produce slow-varying sparse components; at the same time, stochastic sources, such as friction, turbulence and cavitation, produce broadband noise.

The past thirty years have witnessed the development of powerful new techniques for analysing sparse speech and music signals. These operate on the fine structure of narrowband signals, rather than the distribution of signal power. We shall refer to these methods generally as *temporal representations*. Temporal representations include reassigned spectrograms (Kodera *et al.*, 1976, 1978; Gardner and Magnasco, 2006; Fulop and Fitz, 2006), the ensemble interval histogram (EIH Ghitza, 1988; Chandrasekhar and Sreenivas, 2005), zero-crossings with peak amplitudes (ZCPA; Kim *et al.*, 1999; Haque *et al.*, 2007), in-band synchrony (Cooke, 1991/1993; Seneff, 1988), sinusoidal representations (McAuley and Quatieri, 1986), and fine-structure spectrography (Dajani *et al.*, 2005). The mammalian ear itself also belongs to this class of system (Pickles, 2012), and it can be modelled as a cochlear filtering stage followed by non-linear transforms on the fine structure in band-pass signals (e.g., Sumner *et al.*, 2003). Signal transforms that rely on fine structure in the full band include the analytical zero crossing-based methods of Kay and Sudhaker (1986) and Kumaresan and Wang (2001).

Although passive sonar analysis is effectively a machine-listening problem, biologically-motivated signal transforms based on zero crossings (ZC) and instantaneous frequency (IF) have failed to make the same impact in sonar as they have in speech and music analysis. Instead, state-of-the-art processing in sonar continues to rely on linear filtering methods combined with statistical tests performed on measurements of signal power (Burdic, 2003).

The reasons for this are not difficult to discern. Reassigned representations of clean, sparse signals (e.g., a speech utterance recorded in quiet conditions; Gardner and Magnasco, 2006, Fig. 9) possess a visually-impressive, sharp definition, when contrasted with their Fourier-based counterparts. Sonar tonal components, though sparse, are recorded at far lower signal-to-noise ratios (SNRs), and consequently the gains associated with precise component resolution are offset by the blurring effects of noise (rather than those of windowing). More significantly, the mathematical tractability required to quantify the performance of a sonar statistically (McDonough and Whalen, 1995) is lost whenever temporal representations introduce nonlinear operations such as the measurement of zero crossings and instantaneous frequency.

The purpose of this article is to reconcile the benefits of temporal representations on the one hand, with the stringent requirements of sonar on the other. In so doing, we explore two core areas in sonar: detection (Section III) and display (Section IV).

A. Optimal Signal Detection using Temporal Features

The classical, power-based approach to sonar signal detection involves a linear filtering stage that retains signal power and remove noise power, followed by the comparison of a power measurement (or the average of several measurements) to a threshold to optimally decide whether a signal is present. The performance of a power detector can be characterised analytically for familiar classes of stationary signal, such as Gaussian noise and sine waves (McDonough and Whalen, 1995).

Here, we investigate simple temporal counterparts of the power detector, and attempt to detect a target using a single measurement of a zero crossing interval or the instantaneous frequency. A similar project was undertaken fifty years ago by Rainal (1966, 1967), who devised a detector based on the statistics of a zero crossing counter to detect radar signals in clutter (see also Bom and Conoly, 1970; Higgins, 1980). Revisiting this line of research, we find that the performance of a single interval detector compares favourably with a power

detector in a range of theoretical circumstances, especially those in which the target signal is displaced from the centre of the analysis filter. We also derive a joint interval-peak detector, which makes optimal decisions on the basis of a single measurement of a zero crossing interval and its peak square amplitude, thus combining the benefits of the power detector and interval detector.

B. Signal Displays based on Temporal Features

Power-based sonar displays consist of a normalised short-time discrete Fourier transform (DFT). The signal power is divided at regular intervals into frequency bins of equal width and displayed as an image. The resolution of the pixels in this image depend on the number of samples in the DFT: fewer samples provide better temporal resolution at the expense of poorer frequency resolution, and vice versa. The linearity of the DFT allows the means and variances of the DFT magnitudes to be derived analytically for any wide-sense stationary random process with a known autocovariance function, via the Wiener-Khinchine theorem. In particular, if the process is described by sinusoids in Gaussian noise, then the full analytical distribution of the magnitudes is available (Rice, 1944).

Temporal representations reassign the energy in sparse signals according to the fine structure of its constituent components, and consequently provide a sharper time-frequency image, especially in regions where the SNR is high. Nevertheless, the theoretical intuition as to how noisy signals appear in temporal time-frequency representations lags well behind that for the DFT making it an unattractive candidate for sonar purposes. Our contribution in this paper closes this gap somewhat by focussing on two issues. The first is to derive a method for moving from a statistical characterisation of the input random process, e.g., its autocorrelation function or power spectral density, to its temporal pseudo-spectrum. The second is to re-examine the assumptions underlying how fine structure is remapped to frequency when the SNR is poor. Typically, for instance, temporal representations map an interval of i seconds between successive crossings in opposite directions (or some equivalent

measurement) to a frequency estimate $f = 1/(2i)$ Hz, on the assumption that the estimate is unbiased. This assumption is invalid, however, especially at low SNRs. Consequently, inferences about sparse signal components based on fine structure measurements must be made cautiously. This leads to the notion of a *double reassignment* of signal energy: firstly on the basis of a measurement of temporal fine structure, and secondly on the basis of how noise is likely to have corrupted that measurement.

The sections on detection and display draw on some common statistical results, which we derive in Section II: namely, the probability density function (p.d.f.) and cumulative distribution function (c.d.f.) of zero crossing intervals and of the instantaneous frequency.

II. STATISTICAL DISTRIBUTIONS

Detection and estimation procedures decide which of a set of random signals gave rise to a test statistic on the basis of the conditional distributions of those statistics under various hypotheses. In this section, we derive the distributions utilised in the applications and examples that follow in later sections.

We consider the following three test statistics based on temporal features of the signal: (i) the time interval between two successive zero crossings, (ii) a single sample of the instantaneous phase, and (iii) a bivariate statistics consisting of a zero crossing interval paired with its peak square amplitude.

We restrict our hypotheses to random processes that consist of a sinusoid with uniformly random phase added to a zero-mean, wide-sense stationary Gaussian noise process. We routinely start by deriving a discrete-time process with sample values x_n , autocovariances $\gamma_k \triangleq E\{x_n x_{n-k}\}$ and autocorrelation coefficients $\rho_k \triangleq \gamma_k/\gamma_0$. The sampling interval (in seconds) is denoted Δt . Where possible, we derive continuous-time counterparts, $x(t)$, $\gamma(\tau)$ and $\rho(\tau)$, by taking a limit as the sampling interval approaches zero.

A. Zero Crossing Rate

A *zero crossing* in a random process can be defined as a random event in which there is a change of sign between two consecutive samples.

The probability that one sample, x_n , of a zero mean Gaussian process is positive is $\frac{1}{2}$. If the process is also wide-sense stationary, the probability that two samples, x_n and x_{n-k} , are both positive is then a function of the autocorrelation coefficient ρ_k alone (Kedem, 1986):

$$\begin{aligned}\Phi_k &\triangleq \Pr(x_n \geq 0, x_{n-k} \geq 0) \\ &= \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_k.\end{aligned}\tag{1}$$

The probability that three samples, x_n , x_{n-j} and x_{n-k} , are all positive is a sum of bivariate orthant probabilities (David, 1953):

$$\begin{aligned}\Pr(x_{n_1} \geq 0, x_{n_2} \geq 0, x_{n_3} \geq 0) \\ = \frac{\Phi_{n-j} + \Phi_{n-k} + \Phi_{k-j}}{2} - \frac{1}{4}.\end{aligned}\tag{2}$$

In discrete time, a zero crossing occurs at sample n , if the sample at $n - 1$ has the opposite sign. For a stationary Gaussian process, a zero crossing is a random event with time-invariant probability

$$\begin{aligned}\Pr(x_n \geq 0, x_{n-k} < 0 \text{ or } x_n < 0, x_{n-k} \geq 0) \\ = 1 - 2\Phi_1 \\ = \frac{1}{2} - \frac{1}{\pi} \arcsin \rho_k.\end{aligned}\tag{3}$$

The expected zero crossing rate for the discrete-time process x_n , in crossings per second, is therefore

$$\lambda\{x_n\} \triangleq \frac{\pi - 2 \arcsin \rho_k}{2\pi\Delta t},\tag{4}$$

and the expected zero crossing rate for a continuous-time process, $x(t)$, is subsequently obtained by taking the limit of this expression as the sampling interval goes to zero:

$$\begin{aligned}\lambda\{x(t)\} &= \lim_{\Delta t \rightarrow 0} \left\{ \frac{\pi - 2 \arcsin \rho(\Delta t)}{2\pi\Delta t} \right\} \\ &= \frac{1}{\pi} \sqrt{-\rho''(0)}.\end{aligned}\tag{5}$$

This formula, due to Rice (1944), gives the average zero crossing rate purely as a function of the second derivative of the autocorrelation function evaluated at $\tau = 0$.

B. Zero Crossing Intervals

Similarly, a *zero crossing interval* in a random process can be defined as a random event in which there is a change of sign between two consecutive samples, followed by a reversion of sign later on.

We now derive an approximation for the probability distribution governing a single zero crossing interval in either a discrete or a continuous-time process. Consider the probability that the sample x_{n-k-1} is negative, given that a zero crossing from positive to negative occurs at time n . We can rewrite this conditional probability, first in terms of pairwise orthant probabilities, and then in terms of the arcsine of the autocorrelation coefficient using (1) and (2), as follows:

$$\begin{aligned}
& \Pr(x_{n-k-1} < 0 \mid x_{n-1} \geq 0, x_n < 0) \\
&= \frac{\Pr(x_{n-k-1} < 0, x_{n-1} \geq 0, x_n < 0)}{\Pr(x_{n-1} \geq 0, x_n < 0)} \\
&= \frac{1 + 2\Phi_{k+1} - 2\Phi_k - 2\Phi_1}{2 - 4\Phi_1} \\
&= \frac{1}{2} + \frac{\arcsin \rho_{k+1} - \arcsin \rho_k}{\pi - 2 \arcsin \rho_1}. \tag{6}
\end{aligned}$$

The observation of a zero crossing at time n coupled with a reversion of sign at time $n - k - 1$ implies that *at least one* unseen zero crossing falls between times $n - k$ and n . This in turn implies that the interval preceding the crossing cannot exceed k samples in length. However, the converse proposition—that an interval shorter than or equal to k implies the pattern of sign changes in (6) (or its complement)—does *not* hold, as there remains the possibility of multiple crossings.

Let us now assume that the ZC intervals in the process are strictly confined to an octave range, $k_0 < k < 2k_0$. For instance, the intervals may be conditioned in this way using a linear filter. In this case, intervals shorter than or equal to k_0 are impossible, and intervals

shorter than or equal to $2k_0$ are certain. These restrictions permit one to interpret the isolated probabilities in (6) collectively as the cumulative distribution function (c.d.f.) of a random variable, K :

$$\Pr(K \leq k) = \begin{cases} 0 & k \leq k_0 \\ 1 & k \geq 2k_0 \\ \frac{1}{2} + \frac{\arcsin \rho_{k+1} - \arcsin \rho_k}{\pi - 2 \arcsin \rho_1} & \text{otherwise.} \end{cases}$$

This result can be readily extended to continuous-time processes. Let I denote the random variable governing the zero crossing intervals of $x(t)$. Setting $I = K\Delta t$ and using a limiting approach similar to that used in (5), the cumulative distribution function (c.d.f.) for I is

$$\Pr(I \leq i) = \begin{cases} 0 & i \leq i_0 \\ 1 & i \geq 2i_0 \\ \frac{1}{2} + \frac{\rho'(i)}{2\sqrt{\rho''(0)[\rho^2(i) - 1]}} & \text{otherwise,} \end{cases} \quad (7)$$

and the probability density function (p.d.f.) is obtained by differentiation,

$$p_I(i) = \begin{cases} \frac{\rho''(i)[\rho(i)^2 - 1] - \rho(i)[\rho'(i)]^2}{2\sqrt{\rho''(0)[\rho^2(i) - 1]}^3} & 0 \leq i \leq 2i_0 \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

Here ρ' and ρ'' denote the first and second derivatives of the autocorrelation coefficient $\rho(\tau)$ with respect to τ .

The analytical expressions for the c.d.f. and p.d.f. of the zero crossing intervals have been verified numerically for an example random process. The example process is formed by passing white noise with unit power per 1 Hz band through a filter whose squared magnitude response is plotted in Figure 1A. The autocorrelation coefficient of the post-filter process is plotted in Figure 1B. The c.d.f. and p.d.f. of the zero crossing intervals are plotted in Figures 1C and 1D, respectively. The analytical and empirical distributions agree closely in this example, as they do for other test processes (results not shown), so we shall proceed with the analytical result obtained in (8).

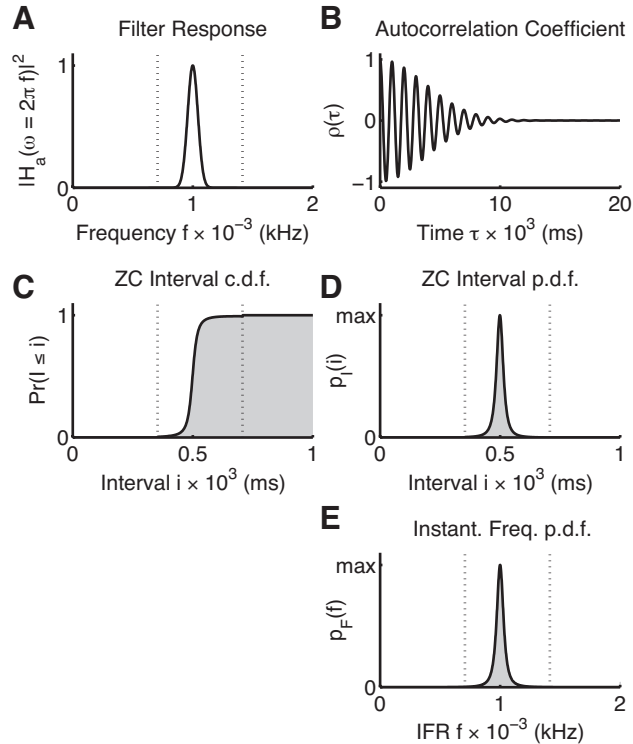


FIG. 1. Analytical and empirical distributions for the zero crossing intervals and instantaneous frequency samples of a Gaussian random process formed by passing white noise through a linear filter. A) Squared magnitude response of the filter used to generate the random process (see text). B) Autocorrelation coefficient of the resultant process. C) Cumulative distribution function of a single zero crossing interval. D) Probability density function of a single zero crossing interval. E) Probability density function of an instantaneous frequency measurement taken from two samples separated by $\Delta t = 2^{-15}$ s. In panels C–E, the analytical approximations are shown as black curves; the empirical distributions, sampled from a 300 s random signal, are shown as a grey areas. The dotted vertical lines indicate frequencies or intervals ± 0.5 octave either side of the centre frequency of the process (1 kHz).

C. Instantaneous Frequency

We now derive a probability distribution to govern the instantaneous frequency for a discrete-time process (see also Angelsen, 1981, for a distinct derivation.) The IF is measured

by taking two consecutive samples of the signal phase, Φ_{n-1} and Φ_n , and calculating the rate of phase change, i.e.,

$$F = \frac{\Phi_n - \Phi_{n-1}}{2\pi\Delta t}, \text{ Hz.}$$

Consequently, the distribution for F can be obtained by considering the distribution of the difference between two successive phase samples, $\phi = \Phi_2 - \Phi_1$.

The p.d.f. governing two consecutive samples in the circular complex, discrete-time process, $z_n = x_n + j\hat{x}_n$, where \hat{x} denotes the discrete Hilbert transform of x , is

$$p(x_{n-1}, \hat{x}_{n-1}, x_n, \hat{x}_n) = \frac{1}{(2\pi)^2 |\Sigma|^{\frac{1}{2}}} \exp\left(\frac{\mathbf{z}^T \Sigma^{-1} \mathbf{z}}{-2}\right),$$

in which

$$\mathbf{z} \equiv \left(x_{n-1}, \hat{x}_{n-1}, x_n, \hat{x}_n \right)^T$$

$$\Sigma \equiv \begin{bmatrix} 1 & 0 & \rho_1 & \hat{\rho}_1 \\ 0 & 1 & -\hat{\rho}_1 & \rho_1 \\ \rho_1 & -\hat{\rho}_1 & 1 & 0 \\ \hat{\rho}_1 & \rho_1 & 0 & 1 \end{bmatrix},$$

and $\hat{\rho}_k$ is the discrete-time Hilbert transform of ρ_k .

Performing the change of variables,

$$\begin{aligned} x_1 &= r_1 \cos \theta, & y_1 &= r_1 \sin \theta, \\ x_2 &= r_2 \cos(\theta + \phi), & y_2 &= r_2 \sin(\theta + \phi); \end{aligned}$$

and marginalising θ , one arrives at the joint probability density for the linear magnitudes of the two samples, r_1 and r_2 , and their phase difference ϕ ,

$$\begin{aligned} &p(r_1, r_2, \phi) \\ &= \frac{r_1 r_2}{2\pi(1 - |\eta_1|^2)} \\ &\quad \times \exp\left[\frac{r_1^2 + r_2^2 - 2r_1 r_2(\rho_1 \cos \phi + \hat{\rho}_1 \sin \phi)}{-2(1 - |\eta_1|^2)}\right], \end{aligned}$$

where $\eta_k = \rho_k + j\hat{\rho}_k$.

Now substitute the two magnitudes for $r_1 = \Gamma \cos \psi$ and $r_2 = \Gamma \sin \psi$, to yield

$$p(\Gamma, \phi) = \int_0^{\frac{\pi}{2}} \frac{\Gamma^3 \sin(2\psi)}{4\pi(1 - |\eta_1|^2)} \times \exp \left\{ \frac{\Gamma^2 [1 - \sin(2\psi)|\eta_1| \cos(\phi - \angle\eta_1)]}{-2(1 - |\eta_1|^2)} \right\} d\psi$$

and then integrate to obtain the marginal distribution for ϕ ,

$$\begin{aligned} p(\phi) &= \int_0^{\frac{\pi}{2}} \int_0^{\infty} \frac{\Gamma^3 \sin(2\psi)}{4\pi(1 - |\eta_1|^2)} \times \exp \left\{ \frac{\Gamma^2 [1 - \sin(2\psi)|\eta_1| \cos(\phi - \angle\eta_1)]}{-2(1 - |\eta_1|^2)} \right\} d\Gamma d\psi \\ &= \frac{1 - |\eta_1|^2}{2\pi} \int_0^{\frac{\pi}{2}} \frac{\sin(2\psi) d\psi}{[1 - |\eta_1| \cos(\phi - \angle\eta_1) \sin(2\psi)]^2} \\ &= \frac{1 - |\eta_1|^2}{2\pi} \left[\frac{q_2(\arctan q_2 + \frac{\pi}{2}) + 1}{1 - q_1^2} \right], \end{aligned}$$

where

$$\begin{aligned} q_1 &\equiv |\eta_1| \cos(\phi - \angle\eta_1) \\ q_2 &\equiv \frac{q_1}{\sqrt{1 - q_1^2}}. \end{aligned}$$

Ultimately then, the distribution for the IF is given by

$$p_F(f) = \frac{p(\phi)}{2\pi\Delta t}. \quad (9)$$

Note that this quantity depends on the sampling interval, Δt . The IF distribution for a continuous signal cannot be obtained in the usual way (by allowing $\Delta t \rightarrow 0$), because $p(\phi)/\Delta t \rightarrow 0$: IF measurements sampled over shorter time periods are noisier, such that the IF distribution approaches infinite variance as the sampling interval goes to zero. Consequently, the IF measurements reported in this article are always with reference to some baseline sampling interval.

Figure 1E plots the analytical IF distribution for an example random process, alongside an empirical version randomly sampled with $\Delta t = 2^{-15}$ s. The derivation for the IF distribution involves no approximations, and consequently, any discrepancy between the analytical and empirical results are due only to the finite sample size (300 s).

D. Zero Crossing Interval and its Peak Square Amplitude

The distributions derived in the preceding paragraphs dealt solely with amplitude scale-invariant features of the signal, such as zero crossings and IF. In this section, we consider the joint density of the zero crossing interval and its peak square amplitude. As an approximation of the peak amplitude, we trust that the signal is suitably narrowband, such that the peaks occur halfway between the crossings, and proceed from there (see Figure 2A).

Assume that a zero crossing interval of duration i has occurred. Let ξ_1 and ξ_2 denote the amplitudes near the crossings, and x denote the amplitude at the midpoint (see Figure 2A). In a stationary Gaussian process, the p.d.f. governing the three amplitudes is

$$p(\xi_1, x, \xi_2) = \frac{1}{(2\pi)^2 |\Sigma|^{\frac{1}{2}}} \exp\left(\frac{\mathbf{z}^T \Sigma^{-1} \mathbf{z}}{-2}\right), \quad (10)$$

in which

$$\mathbf{z} \equiv \left(\xi_1, x, \xi_2 \right)^T$$

$$\Sigma \equiv \begin{bmatrix} \gamma(0) & \gamma(\frac{i}{2}) & \gamma(i) \\ \gamma(\frac{i}{2}) & \gamma(0) & \gamma(\frac{i}{2}) \\ \gamma(i) & \gamma(\frac{i}{2}) & \gamma(0) \end{bmatrix}.$$

Conditioning this p.d.f. upon $\xi_1 = \xi_2 = 0$ to mimic the effect of zero crossings is incorrect. Interpreted explicitly, (10) gives the probability mass residing in the differential element $[\xi_1, \xi_1 + \Delta\xi_1]$, $[\xi_2, \xi_2 + \Delta\xi_2]$, $[x, x + \Delta x]$, divided by its volume, $\Delta\xi_1 \Delta\xi_2 \Delta x$. In contrast we aim to find the density in the region $[x + \Delta x]$, $[i + \Delta i]$. In order to relate the two, we assume that $x(t)$ is a narrowband process, whose sample functions resemble sinusoids on short time scales. We can therefore safely substitute the waveform around the peak, x , for a cosine, and then approximate the differential change Δi as arising from the differential changes $\Delta\xi_1, \Delta\xi_2$, as depicted in Figure 2A.

The change of variables

$$x = \sqrt{a}$$

$$\Delta i_1 = (+\pi\sqrt{a}/i)\Delta\xi_1$$

$$\Delta i_2 = (-\pi\sqrt{a}/i)\Delta\xi_2,$$

converts the p.d.f. in (10) into one that governs the square of the midpoint ($a = x^2$) and the *temporal* displacement of the zeros ($\Delta i_1, \Delta i_2$).

Now, having adjusted differential areas, conditioning on $\Delta i_1 = \Delta i_2 = 0$ results in the distribution of the squared midpoint,

$$p_{A|I}(a | i) = \sqrt{\frac{a}{2\pi\beta^3}} \exp\left(\frac{a}{-2\beta}\right), \quad (11)$$

where

$$\beta \equiv \frac{\gamma(i) - 2\gamma(\frac{i}{2})\rho(\frac{i}{2}) + \gamma(0)}{1 + \rho(i)}.$$

The joint distribution for the square peak A and the zero crossing interval I is then

$$p_{IA}(i, a) \equiv p_{A|I}(a | i)p_I(i). \quad (12)$$

Figure 2B plots a few contours of the joint p.d.f. for the zero crossing intervals and squared peaks of the process described in Figure 1A. The close agreement between the empirical (grey) and analytical (black) contours supports the preceding work in this section. The quality of the contours also resemble those once found by Longuet-Higgins (1983) using an alternative analytical approach. The marginal density functions, $p_I(i)$ and $p_A(a)$, found by integrating the joint p.d.f. (numerically), also closely agree..

E. A Randomly-Phased Sinusoid in Noise

In the preceding sections the statistical distributions of some timing-based features of Gaussian processes were derived. Another rudimentary class of random process, besides Gaussian, are sinusoids, or processes consisting of a mixture of a sinusoid and noise. A pure sinusoid with uniformly random phase, fixed amplitude α , and fixed radial frequency ω , is

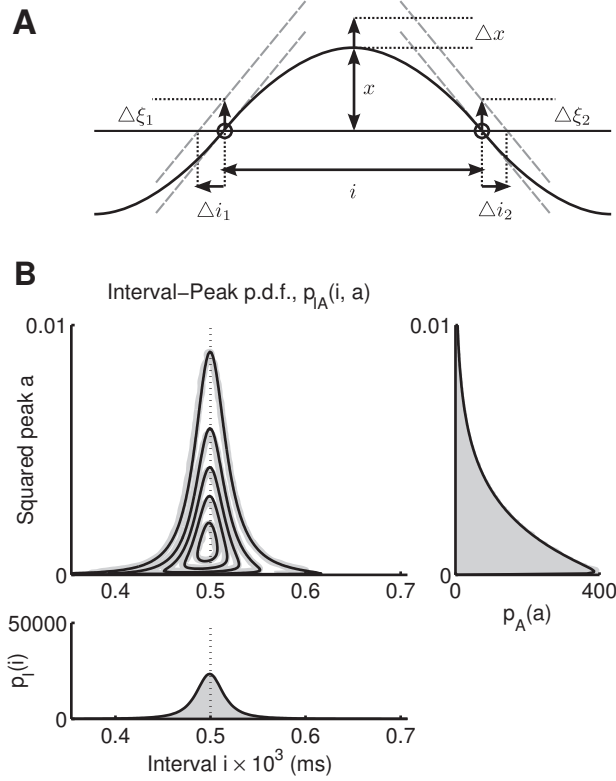


FIG. 2. Joint statistics of intervals and peaks. A) Diagram illustrating how small changes to sample values around zero, $\Delta\xi_1$ and $\Delta\xi_2$, relate to small changes in crossing interval, Δi_1 and Δi_2 , provided that one assumes a functional form for the local waveform. B) Five contours of the joint interval-peak p.d.f., $p_{IA}(i, a)$, uniformly spaced between zero and its maximum. The thick grey lines are contours of the empirical distribution; the thinner black lines are those of the analytical distribution. Marginals for $p_I(i)$ and $p_A(a)$ are shown to the bottom and right, respectively. The p.d.f. is that of the example Gaussian process used in Figure 1.

a non-Gaussian, wide-sense stationary random process with the following autocovariance function (McDonough and Whalen, 1995):

$$\gamma_s(\tau; \mathcal{A} = \alpha) = \frac{\alpha^2}{2} \cos \omega\tau.$$

Because this process is non-Gaussian, we cannot employ the formulae above (i.e., those for intervals and IF) using $\gamma(\tau)$, without violating the Gaussian assumptions that were in place

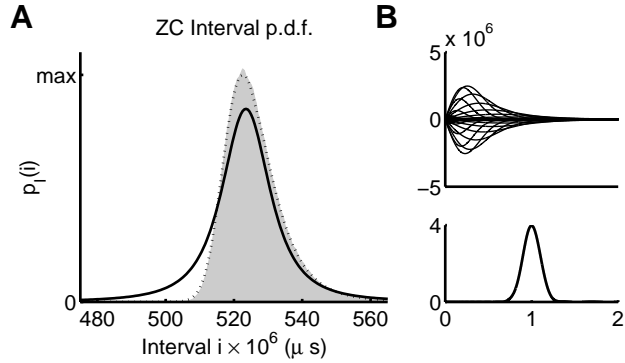


FIG. 3. A) The distribution of the zero crossing intervals of a sinusoid in noise, based on a 300 s sample (grey area). The sinusoid has frequency 950 Hz and amplitude $\alpha \approx 127$, and the noise is white with unit power per 1 Hz band, giving a narrow-band SNR of ≈ 33 dB. The ZC intervals were measured after the mixture was passed through a filter with the magnitude response plotted in Figure 1A (right).

during their derivations. If we proceed with the formula for the zero crossing intervals in (8) anyway, the result substantially differs from the true distribution (Figure 3A). Any departure from the true distribution is detrimental to detection and estimation performance. In this section the solution we have already obtained is adapted so that it works for a sinusoid in Gaussian noise.

A closely related process, which *is* Gaussian, yet has sinusoidal sample functions, is a sinusoid with uniformly random phase and a random (but constant) amplitude, drawn initially from a Rayleigh distribution with scale parameter σ . Its autocovariance function is

$$\gamma_s(\tau; \mathcal{A} \sim \text{Rayl}\{\sigma\}) = \sigma^2 \cos \omega \tau.$$

Using a Dirac delta function, a fixed amplitude can be interpreted as a random variable, whose probability mass is concentrated at a single point. Let us then assume that one can approximate a Dirac pulse as a superposition of Rayleigh density functions at different

scales. Combining these two ideas, we can write

$$\begin{aligned}
p_I(i; \mathcal{A} = \alpha) &= \int_0^\infty p_I(i; \alpha) \delta(\alpha' - \alpha) d\alpha' \\
&= \int_0^\infty p_I(i; \alpha) \left[\int_0^\infty \eta(\sigma) p(\alpha; \sigma) d\sigma \right] d\alpha' \\
&= \int_0^\infty \eta(\sigma) \left[\int_0^\infty p_I(i; \alpha) p(\alpha; \sigma) d\alpha' \right] d\sigma \\
&= \int_0^\infty \eta(\sigma) p_I(i; \mathcal{A} \sim \text{Rayl}\{\sigma\}) d\sigma,
\end{aligned} \tag{13}$$

where $p(\alpha; \sigma)$ is the p.d.f. of a Rayleigh distribution:

$$p(\alpha; \sigma) = \frac{\alpha}{\sigma^2} \exp\left(-\frac{\alpha^2}{2\sigma^2}\right).$$

Notice that the densities $p_I(i; \mathcal{A} \sim \text{Rayl}\{\sigma\})$ are *known* quantities: namely, densities of the distributions of ZC intervals of band-limited, stationary Gaussian processes, an analytical formula for which was provided in (8).

The sole remaining task is to find a function $\eta(\sigma)$ that satisfies the inner product equation

$$\begin{aligned}
\int_0^\infty \eta(\sigma) p(\alpha; \sigma) d\sigma &= \int_0^\infty \eta(\sigma) \frac{\alpha'}{\sigma^2} \exp\left(\frac{-\alpha'^2}{2\sigma^2}\right) d\sigma \\
&= \delta(\alpha' - \alpha).
\end{aligned} \tag{14}$$

It is unlikely that there exists a function that meets the requirement in (14) analytically. However, a numerical approximation can be obtained by minimising the total squared error between a linear combination of a finite number of sampled Rayleigh densities and a (finitely) narrow pulse. (We describe this procedure in the appendix.)

Figure 3 demonstrates how a sum of 25 appropriately weighted Rayleigh densities (Figure 3B) can produce a narrow pulse (Figure 3C). Zero crossing interval p.d.f.s for Gaussian processes are then combined linearly using these weights, and a much better approximation to the true distribution results (Figure 3A: dotted line). This procedure is quite general, applying in any situation that demands a result for the sum of a sinusoid and a Gaussian process, when one possesses only the solution for a purely Gaussian random process. For instance, it can be used to approximate the distribution of the instantaneous frequency of a sinusoid in noise, from the noise-only solution. (We do not present an exemplar, but utilise

the result in Section III.) An alternative approach to generating this density is described in Cobb (1965).

III. OPTIMAL SIGNAL DETECTION

The statistical distributions derived in the previous section can be incorporated readily into Bayes' rule to form the core of timing-based detectors. These detectors perform optimally, in the sense that no other detector which employs the test statistic in question (e.g., a zero crossing interval) is superior. Formally, such detectors consist of a rule that compares a likelihood ratio formed from the posterior probabilities of the measurement for two hypotheses (the signal-and-noise hypothesis, H_1 , and the noise-only hypothesis, H_0) to a likelihood threshold, λ .

A. Hypothesis Tests

We now compare the performance of four detectors in a sine-in-noise detection task. The *zero crossing interval detector* measures a single zero crossing interval, i , and utilises the decision rule

$$\text{Choose } H_1 \text{ if } \frac{p_I(i | H_1)}{p_I(i | H_0)} \geq \lambda,$$

and H_0 otherwise.

Here, $p_I(i | H_0)$ is the p.d.f. of the zero crossing intervals of a Gaussian process (8), and $p_I(i | H_1)$ is the p.d.f. of the zero crossing intervals of a sine wave added to a Gaussian process (13).

The *instantaneous frequency detector* measures a single sample of the instantaneous frequency, f , and utilises the decision rule

$$\text{Choose } H_1 \text{ if } \frac{p_F(f | H_1)}{p_F(f | H_0)} \geq \lambda,$$

and H_0 otherwise.

Here, $p_F(f | H_0)$ is the probability density function for the instantaneous frequency of a Gaussian process (9), and $p_F(f | H_1)$ is the p.d.f. of the instantaneous frequency for a sine

wave added to a Gaussian process, which is derived according to the technique set out in Section II.E.

The *joint interval-peak detector* measures a single zero crossing interval, i , and its peak squared amplitude, a , and utilitises the decision rule

$$\text{Choose } H_1 \text{ if } \frac{p_{IA}(i, a | H_1)}{p_{IA}(i, a | H_0)} \geq \lambda,$$

and H_0 otherwise.

Here, $p_{IA}(i, a | H_0)$ is the joint p.d.f. governing the zero crossing interval and its peak squared amplitude for a Gaussian process (12), and $p_{IA}(i, a | H_1)$ is the joint p.d.f. for a sine wave added to a Gaussian process, also derived using the technique described in Section II.E.

The *squared-envelope (quadrature) detector* measures a single sample of the signal envelope, $|Z|$, and utilises the decision rule

$$\text{Choose } H_1 \text{ if } I_0\left(\sqrt{\frac{S}{N}}|Z|\right) \geq \lambda,$$

and H_0 otherwise.

where $\frac{S}{N}$ is the ratio of signal power to noise power at the output of the analysis filter, and I_0 denotes a zeroth-order modified Bessel function of the first kind.

B. Minimum-Error Criterion

We first compare the performance of the four detectors for a minimum error detection task, where the probability of a signal is 50%. This is achieved by setting $\lambda = 1$ (McDonough and Whalen, 1995). Detection performance is measured as the probability of a correct decision, and is determined analytically and using simulations.

Sinusoidal signals are presented against a white noise background with various narrow-band SNRs (NB-SNR). The NB-SNR refers to the ratio of the total signal power to the noise power in a 1 Hz bandwidth. (Consequently, the NB-SNR excludes the effects of pre-filtering.) We also define two fixed SNRs: a “low” NB-SNR of 20 dB and a “high” NB-SNR of 40 dB.

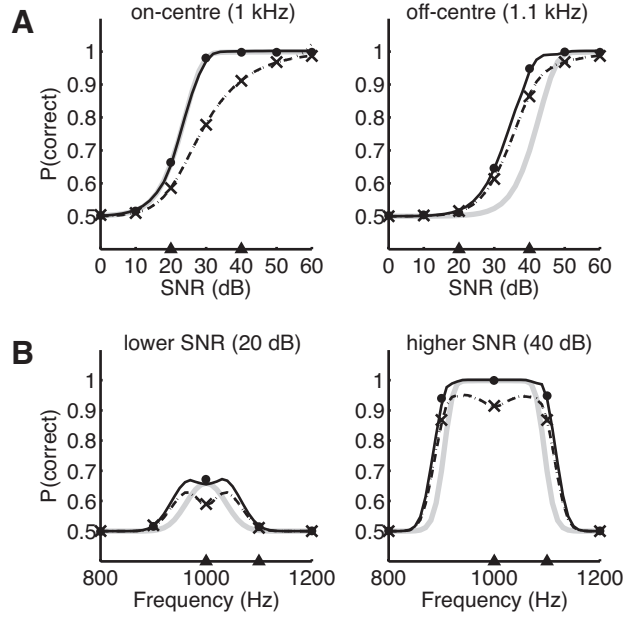


FIG. 4. Performance of the squared-envelope (thick, grey line), interval (dashed line; crosses), instantaneous frequency (dotted) and joint interval-peak (solid line; round marker) detectors, when following a minimum error criterion. The analytical and simulations results are plotted using lines and markers, respectively. No empirical performance is measured for the envelope or IF detectors. A) Detection performance as a function of SNR for two signal frequencies (left: on-centre; right: off-centre). B) Detection performance as a function of frequency for two SNRs (left: lower SNR; right: higher SNR).

The signal and noise are received through a linear filter whose magnitude response possesses a Gaussian profile, is centred at 1 kHz, and has a 3 dB bandwidth of 160 Hz. This magnitude response is somewhat wider than that depicted in Figure 1A. The signal is a pure sinusoid of known, constant frequency. We also define two fixed frequencies: “on-centre”, at 1 kHz; and “off-centre” at 1.1 Hz.

Figure 4A plots the probability of a correct decision as the SNR is varied when the signal is on-centre (left panel) and off-centre (right panel). The performance of the four detectors increases monotonically with NB-SNR, and the analytical results (curves) are supported by the simulation results (markers).

In the on-centre case, the detectors that rely on power (the squared-envelope and joint-interval peak detectors) consistently outperform those that rely exclusively on temporal statistics (the ZCI and IF detectors). The power detectors perform similarly, suggesting that the temporal statistics provides no additional information when the signal is centred on the band. The ZCI and IF detectors also perform similarly, as both measurements convey local information concerning the rate of change in the signal phase: the former measures the time between two fixed points in the phase, the latter measures the change in phase between two fixed points in time.

In the off-centre case (right panel), where the signal is placed at 1.1 kHz, the pattern changes. At lower SNRs, the temporal detectors commit fewer errors than the squared-envelope detector; and the joint-interval peak detector outperforms all three. The resolved signal power is reduced when the signal is displaced from the centre, so that detection performance worsens overall (the curves shift to the right). However, at low NB-SNRs, the fine timing structure is more informative than the envelope. The performance of the joint interval-peak detector is in all cases superior, indicating that the power and timing information are mutually independent to some degree.

C. Neyman-Pearson Criterion

The Neyman-Pearson criterion relaxes the requirement that errors of both kinds are minimised, and instead maximises the hit probability, $\Pr(\text{Hit})$, subject to a fixed probability of false alarm, $\Pr(\text{FA})$. This involves setting the likelihood threshold λ to satisfy the integral

$$\int_{\mathcal{R}_1(\lambda)} p_I(x_i | H_0) dt = \Pr(\text{FA}), \quad (15)$$

where \mathcal{R}_1 is the set of zero crossing intervals that leads to a decision in favour of H_1 , given λ .

In practice (15) is difficult to solve for a specific λ . The receiver operating characteristic (ROC) curve plots a general solution to the Neyman-Pearson criterion by varying λ parametrically, and plotting the hit and false alarm probabilities on two-dimensional coordinates.

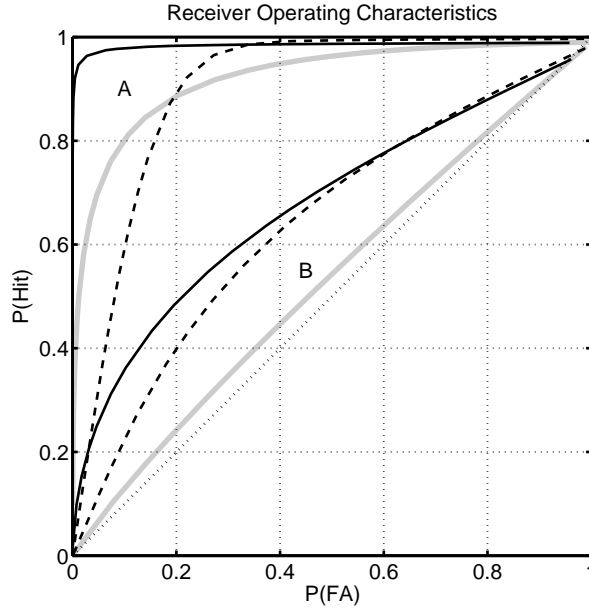


FIG. 5. ROC curves for squared-envelope (thick, grey), interval (dashed) and joint interval-peak (solid) detectors. The analysis band is Gaussian in shape and centred at 1 kHz in all instances. The family of curves marked ‘A’ show the detection performance for a target signal of frequency 1050 Hz. The family of curves marked ‘B’ show detection performance for a target signal of frequency 1100 Hz. In both cases the NB-SNR is 30 dB. Chance performance is marked with a dotted line.

Figure 5 plots ROC curves for the squared-envelope, zero crossing interval and joint interval-peak detectors. (We omit the curve for the IF detector, which coincides with that of the ZCI detector.) We employ the same signal and noise configurations as those described in the previous section. The NB-SNR is set to 30 dB in all conditions. The signal is placed either at 1.05 kHz or 1.1 kHz.

When the displacement from the centre is small (0.07 octaves, curve set A), the curves for the envelope and interval detectors intersect. In a high false alarm regime, the signal is best detected by its influence on the zero crossings; in a low false alarm regime, a sample of the envelope is more informative. At $\Pr(\text{FA}) \approx 0.2$, the two statistics are equally useful. The joint interval-peak detector outperforms the detectors that rely on one type of measurement.

(The apparent improvement of the ZCI detector over the joint detector at high false alarm rates is an artefact arising from error in the approximation of the density functions. See Section II.)

When the displacement from the centre is larger (0.14 octaves, curve set B), all the ROC curves are nearer to the chance performance line, owing to the overall attenuation of the signal as it passes into the tail of the filter. Here the ZCI detector outperforms the squared-envelope detector over the majority of false alarm probabilities shown, with convergence in the corners. This supports the conclusion of the previous section: that displacing the frequency from the band centre influences the zero crossings of the mixture in such way as to render the signal more detectable. Again, a detector that combines information from the envelope and fine structure results in the best performance.

Thus far we have illustrated the superior performance of the ZCI detector using a fairly wide analysis band ($Q = 6.25$). In practice, applications such as narrowband passive sonar simply utilise a discrete Fourier transform with a small analysis bandwidth to achieve a suitably high SNR prior to envelope detection. To see whether the interval and joint interval-peak detectors offer any advantage over the envelope detector at lower analysis bandwidths, we generated variants of the ROC curves shown in Figure 5 by consistently rescaling all quantities by a factor of 0.1 in relation to a fixed band centre of 1 kHz. Thus, the bandwidth was set to 16 Hz, the signal was placed at either 1005 Hz or 1010 Hz (A or B), and the NB-SNR was decreased by 10 dB. There is no visible change to Figure 5 at print resolution following these changes, which suggests that it is the displacement of the signal *in relation to the analysis bandwidth* which determines detectability. The ROC curves are invariant with respect to rescaling by factors other than 0.1 (and smaller than 1; results not shown).

IV. DISPLAY

The previous section described how one can optimally decide between Gaussian processes on the basis of *single* measurements taken from narrowband signals: envelope samples,

zero crossing intervals, instantaneous frequency samples, or zero crossing intervals paired with envelope samples. This section describes how *many* measurements can be taken and compiled to form pseudospectral signal representations.

The pseudospectral representations we now review all correspond in some fashion to the detectors described earlier, as we note in the following section. Because of these common principles, the derivations of Section II can be re-used to approximate the mean spectral profiles analytically. We thereby learn that the representation of the signal in these spectra is biased in the presence of noise. Lastly we demonstrate a proof-of-concept technique that, given sufficient information about the background noise process, counteracts this bias.

A. Algorithms for the Generation of Pseudospectra

Sampling the squared envelope in many narrowband channels and plotting them produces a power spectrum (or an energy spectrum, depending on the normalisation scheme chosen). The discrete Fourier transform falls into this category. These representations relate most closely to an envelope detector, and applying a threshold to a single DFT bin amounts to an implementation of the squared envelope detector.

The ensemble interval histogram (EIH) (Ghitza, 1988), working from abstractions of biological principles found at work in the ear, compiles reciprocal intervals from the output of multi-level crossing detectors in many channels into a histogram. Temporal information is conveyed by the crossing times, and a rudimentary form of envelope information is conveyed by the number of levels crossed. (Energetic signals activate more level crossing detectors.) The pseudospectrum formed from *zero* crossing intervals alone, we refer to as the *zero crossing interval histogram* (ZCIH) (although, properly speaking, the histogram is formed from reciprocal intervals). Evidently, the ZCIH is a specialisation of the EIH, in which the multi-level crossing detector consists solely of the zero level. The ZCIH relies on the same measurement as the zero crossing interval detector.

Modern variants of the reassigned spectrogram (e.g., Gardner and Magnasco, 2006)

consist of a histogram of the instantaneous frequency in each channel (and some temporal adjustments, which we neglect here). These present in a direct fashion the information used by the instantaneous frequency detector.

Finally, the zero crossings with peak amplitudes (ZCPA) algorithm, due to Kim *et al.* (1999), is a weighted histogram of reciprocal zero crossing intervals, in which the individual intervals are weighted by the narrowband envelope, or some function thereof. In its original, biologically-motivated formulation, the intervals were weighted by the log peak amplitude. Adaptations of this algorithm to suit sonar purposes would be likely to use the square of the envelope. This representation bears the closest resemblance to the joint interval-peak detector.

B. Mean Pseudospectra for Gaussian Processes

Figure 6A plots the power spectrum of a Gaussian noise process, on a logarithmic scale. Underneath are plotted the magnitude responses of a bank of filters, on a linear scale. The filter centres are spaced at intervals of 32 Hz and have a fixed bandwidth of 64 Hz. The graphs below (panels B–E) correspond to four (pseudo)spectral representations of 100 hundred seconds of noise. In each case, the grey regions are one empirical measurement, and the black curves (or bars) plot the analytical approximation.

Figure 6B plots the mean squared envelope in each filter channel, both as measured (grey) and as expected analytically (black). In this case, the analytical solution is exact and is based on the mean of the exponential distribution. The resolution of the peaks in the noise profile is limited by the spacing and width of the analysis filters. The sharp resonances in the noise are somewhat narrower than the analysis filters, and, consequently, the peaks are smeared out in the energy representation. The second peak (at 1500 Hz) is particularly obscure.

Figure 6C plots the mean ZCIH. A global histogram is formed from the reciprocal intervals measured in the output of each filter. To be more exact, when a zero crossing

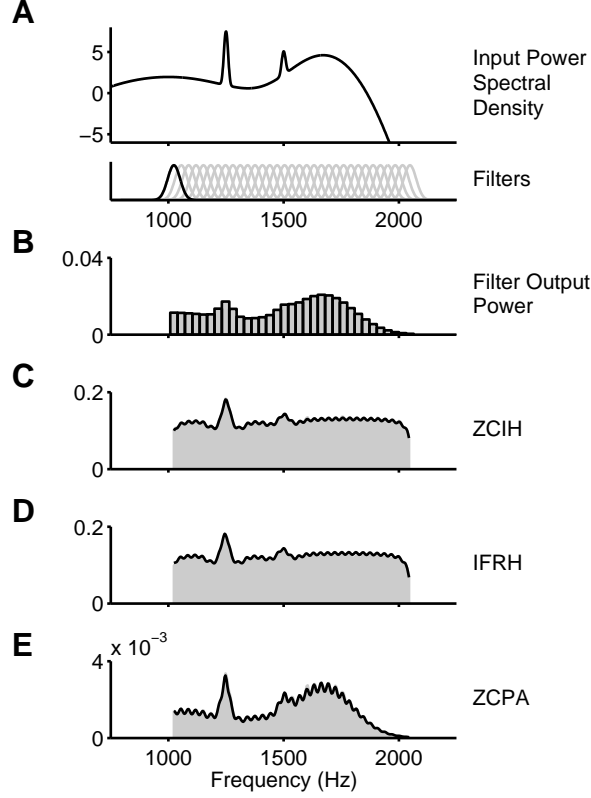


FIG. 6. Pseudospectra based on timing measurements for an example realisation of a Gaussian random process. A) Power spectral density of the process (upper) and squared magnitude frequency responses of the analysis filters (lower). B) Squared magnitude of output samples at each filter. C) Zero crossing interval histogram. D) Instantaneous frequency histogram. E) Zero crossings with peak amplitudes. In panels B–E, the black curves show analytical approximations of the expected profile, and the grey regions show a measurement taken from a random sample of 100 seconds.

interval i is measured in a filter, the histogram bin containing frequency $\frac{1}{2i}$ Hz is incremented. The histogram bins are 4 Hz wide, and the contribution of each interval is weighted by i to adjust for the fact that shorter intervals (corresponding to higher frequencies) occur more often per unit time. The analytical mean profile is generated using the cumulative distribution function of the zero crossing intervals (7). If the lower and upper edges of the histogram bin are f_L and f_U , respectively, then the probability that a channel contributes

to this bin is

$$\Pr(2If_U \leq 1) - \Pr(2If_L < 1),$$

where I is the random variable governing the intervals leaving the channel. The ZCIH is found by summing these probability mass functions over all channels.

The analytical result closely approximates the random sample. Three features of this graph are noteworthy. Firstly, the two prominent peaks in the spectral profile are well represented in the ZCIH. Secondly and relatedly, smooth portions of the spectrum are flat in the ZCIH, because there is no temporal structure presence except for noise. Thirdly, there is a ripple artefact in the ZCIH which originates with the analysis filters. The zero crossing intervals at the output of each narrowband filter are dominated by the periods of the frequencies around the peak. ZCIH ripple can be reduced by increasing the analysis filter bandwidth (at the cost of distinct frequency component resolution) or spacing the filter centres more densely (at the cost of additional computation).

Figure 6D presents the mean instantaneous frequency histogram (IFRH). This representation is derived in the same way as the ZCIH, except that the histogram is formed from samples of the instantaneous frequency (measured at a sampling rate of 32384 Hz), rather than from zero crossing intervals. As we lack an expression for the c.d.f. of the IF distribution, when deriving the analytical result, we use the probability density function at the centre of the histogram bin. The IFRH closely resembles the ZCIH, just as the results for the interval detector follow those of the IF detector. Consequently, the comments above in connection with the ZCIH apply in this case.

Figure 6E displays the mean zero crossings with peak amplitudes (ZCPA). This pseudospectrum is derived in the same manner as the ZCIH, with the exception that the contribution of each zero crossing interval is weighted by the square of its peak amplitude. Consequently, this representation combines aspects of both the ZCIH and the squared magnitude of the Fourier spectrum. Both narrowband “signals” are clearly visible in the ZCPA, and the ratio of the peaks to the surrounding noise floor is at its highest in this form of

spectrum. The analytical result is found by plotting

$$E\{a \mid i\}p_I(i) \equiv \int_0^\infty ap_{IA}(i, a) da.$$

C. Doubly-Reassigned Zero Crossing Interval Histograms

The preceding discussion in relation to detectors and pseudospectra has highlighted that, in the presence of noise, one cannot reliably “read off” the frequency of a signal component from the zero crossings in a band-pass signal. An uneven noise background will cause signal components to gravitate towards energetic regions, and even white noise biases the frequency towards the filter centre. The reassigned spectrograms proposed to date, including those mentioned above, do not account for these factors.

The analytical results derived earlier express the probability density of a zero crossing interval (or IF, peak) in terms of a signal and noise hypothesis: $p_I(i \mid H)$. These densities can be formed into a likelihood ratio that form the core of an effective detector. Here, we deploy Bayes’ rule again, to map zero crossing intervals observed from a filter (which are unreliable) back to true signal frequencies, using a model of the noise floor. In a standard ZCIH, if the interval i is observed, we increment the histogram bin which captures $\frac{1}{2i}$; that is, a single point maps to a single point.

We now propose a variant of the ZCIH in which each observation i contributes the full posterior density to the histogram. Using Bayes’ rule, and keeping the noise background and filter shape implicit, this density is

$$\begin{aligned} p_F(f \mid i; N) &= \frac{p_I(i \mid f)p_F(f)}{p_I(i)} \\ &= \frac{p_I(i \mid f)p_F(f)}{\int p_I(i \mid f')p_F(f') df'}. \end{aligned}$$

This formulation explicitly accounts for prior expectation of where the signal component is to be found. In the absence of such knowledge, we can assume a uniform prior frequency in the range $[f_1, f_2]$ and write

$$p_F(f \mid i; N) = \frac{p_I(i \mid f)}{\int_{f_1}^{f_2} p_I(i \mid f') df'}.$$

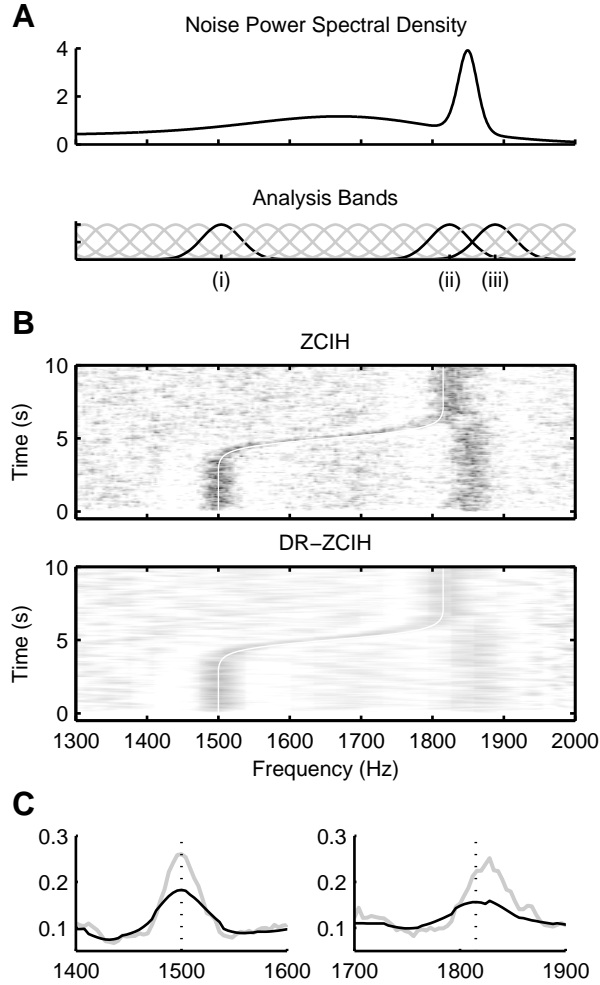


FIG. 7. Example of the doubly-reassigned ZCIH. A) Power spectral density of the noise process (top) and distribution of the analysis filters (bottom). B) ZCIH (top) and DR-ZCIH (bottom). The white line shows the frequency track of a component with amplitude 40. C) Time-averaged ZCIH (grey line) and DR-ZCIH (black line) for the regions 0–3 seconds (left) and 7–10 seconds (right). The dotted line marks the stationary signal frequency.

We refer to this representation as *doubly-reassigned* to reflect this second stage of re-assignment that is made to the timing information to account for the noise. The doubly-reassigned ZCIH we abbreviate DR-ZCIH.

We provide an example of this procedure in Figure 7. Figure 7A plots the power spectral

density of a noise background, with a prominent swathe of noise around 1850 Hz. The analysis filters used by the ZCIH are 64 Hz wide, spaced at intervals of 32 Hz and plotted beneath.

The signal to be displayed is a tonal of ten seconds duration, which consists of a 3-second 1500 Hz segment and a 3-second 1815 Hz segment, connected by a 6-second sigmoidal sweep, plotted as a white line on both panels in Figure 7B. The ZCIH (upper panel) represents the 1500 Hz signal segment with little visible bias. The noise spectrum at this locality is quite smooth. On the other hand, the 1815 Hz signal segment appears smeared in the ZCIH towards the prominent spectral peak in the noise background. The intervals captured by the filters labelled from (ii) to (iii) are influenced by both the signal and the noise.

Prior knowledge of the noise floor permits us to adjust the signal frequency to counteract its effect, as shown in the lower panel of Figure 7. The signal remains centred on the true component at all times and the uneven noise background exerts less influence. The mean ZCIH and DR-ZCIH during the two stationary signal segments are plotted in Figure 7C, making the bias during the 1815 Hz segment in the ZCIH, and its absence in the DR-ZCIH, more apparent. In the DR-ZCIH, the contrast of the signal with the noise background is poorer. This loss of contrast reflects a principled adjustment made for the uncertainty introduced by the noise. So, whilst bumps in the DR-ZCIH may be shallower, their existence and the spread of frequencies they convey are more trustworthy. In the limit of $\text{SNR} \rightarrow -\infty$, the DR-ZCPA is flat; as there is no principled way for it to reassign intervals, so each interval contributes a uniform distribution.

Figure 8 shows how the DR-ZCPA reassigns intervals in three selected filters in this example, marked (i), (ii) and (iii) in Figure 7A. Each vertical column of pixels shows density that is added to the ZCIH for a measured zero crossing interval (x -axis shows $f = \frac{1}{2i}$). The white, diagonal line represents the identity reassignment used by the naïve ZCIH. In the limit of $\text{SNR} \rightarrow +\infty$, the density is concentrated along this line. In the limit of $\text{SNR} \rightarrow -\infty$, the entire image is grey. (Neither of these conditions is shown.) The white curve plots the means of the densities in each column. The black/white dotted lines mark the filter centre

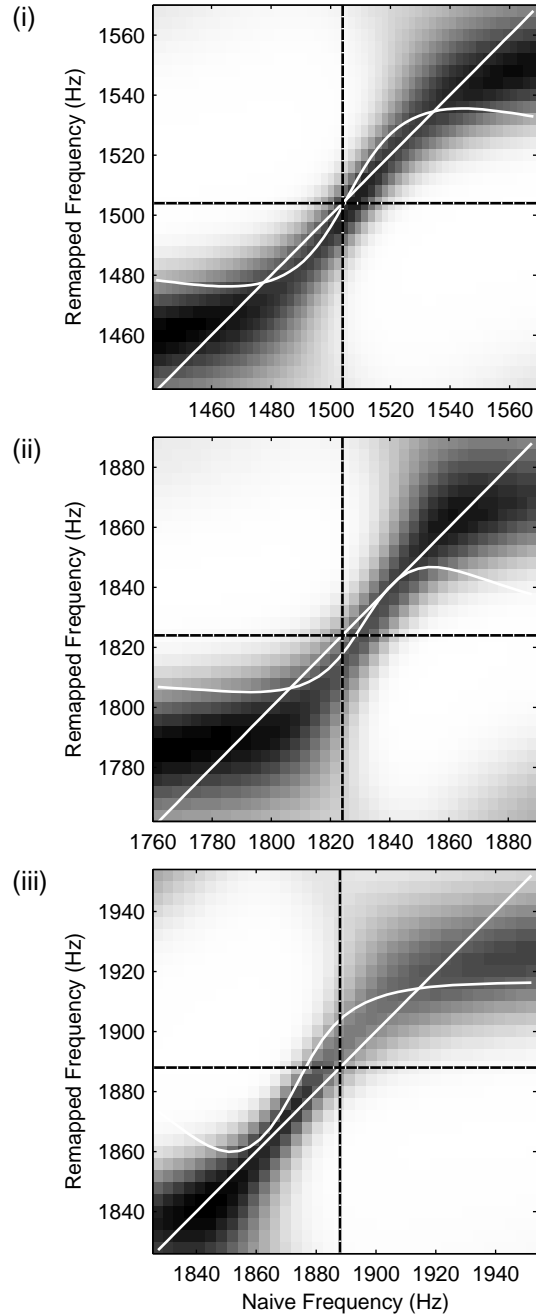


FIG. 8. Frequency reassignment maps used at the three filters marked (i), (ii) and (iii) in Figure 7A. The abscissa shows the measured frequency ($\frac{1}{2i}$) and the ordinate shows the adjusted frequency. Pixel values show the probability density of a frequency adjustment (higher values are darker); vertical pixel columns show posterior distributions. Dotted black/white lines show filter centre frequencies. White diagonal lines plot a naïve (identity) remapping. The S-shaped white lines plot the distribution mean in each column (for visual purposes only, this mean is not used in the DR-ZCIH).

frequency.

Filter (i) is centred on a noise region with a high SNR and a flat spectral profile. Frequencies measurements that coincide with the filter centre are trustworthy, and the four lines converge at the centre. However, even in flat spectral noise conditions, the dominant frequency principle causes the frequency of the output component to gravitate towards the filter centre. The reassignment map therefore mildly extremifies any measurements that fall either side of the filter centre. Measurements that fall in the tail of the filter are unreliable and consequently the reassigned frequencies are most spread out here.

Filters (ii) and (iii) are centred either side of a prominent peak in the noise floor around 1850 Hz. The maps for these filters have an S-shaped appearance similar to that for (i). The noise peak is liable to bias the measurements in filter (ii) upwards; consequently map (ii) adjusts them downwards. Similarly, the peak is also liable to bias the measurements in filter (iii) downwards; consequently map (iii) adjusts them upwards.

V. DISCUSSION

The presence of a signal in an acoustic mixture leads to an excess of energy that renders it detectable. However, it is also possible to detect a signal by its influence on the *structure* of the signal. A clean tonal, for instance, will produce zero crossings too regular to be accounted for by background noise. In this article we have shown that detectors that operate on the fine structure of a signal exceed the performance of those that operate on a sample of the squared envelope when the analysis filter is not ideally configured for power detection. The joint interval-peak detector combines information from the envelope and zero crossings to achieve the best performance.

Although power is likely to be the statistic used for detection in most circumstances, there will be occasions in which timing-based detection is preferable. Examples include situations in which broader analysis bands are used (e.g., to achieve high time resolution or reduce computations), where the gain of the analysis filter is unreliable or unknown, or

where the aim is to detect a signal of a particular frequency in a band, as opposed to any source of energetic noise. The interval detector, being keyed on structure not energy, is able to distinguish genuinely signal-related contributions to the filter output from other sources, such as noise bursts and clutter.

Passive sonar displays convey the power levels in DFT bins as the colour or greyscale of a row of pixels in a waterfall spectrogram. The final detection decision then rests with the human viewer (Grigorakis, 1997). The information used by timing detectors — zero crossings or IF — can also be used to produce time-frequency displays that reassign energy, e.g., ZCPA. The distributions used in the timing detectors can be used to derive the mean pseudospectra for Gaussian noise and sine-in-Gaussian noise processes. The mean profile for white noise can be calculated to highlight undesirable ripples due to frequencies gravitating towards filter centres. The fact that that the ripples resemble tonal components is especially unfortunate, so the ability to check analytically that the mean profile is suitably flat before deployment — by choosing the appropriate number of filters, and their spacing and shape — is valuable.

In addition to the analysis filters spuriously reassigning energy, there is also the problem of signal structure gravitating towards energetic portions in the noise spectrum. The distributions relating the zero crossings or instantaneous frequency measurements to signal frequency can be used again to invert this tendency to some degree, in which case we say the representation is doubly reassigned: the model is used to reassign measurements to counteract the noise background and analysis filters; then the adjusted measurements are in turn used to reassign the energy (or directly plotted in the case of the DR-ZCIH).

It remains to investigate the possible variations upon double reassignment that exist. An interval-peak pair recorded at a linear filter can be viewed as a joint statistic that is used to update a time-frequency display, according to the best hypothesis, or range of hypotheses, that it indicates according to a noise model. Similarly, the work reported here may be usefully applied in “machine hearing” systems that detect and classify sounds according to auditory principles (e.g., Wang and Brown, 2006). Such systems typically include a model

of cochlear filtering in which both firing rate and temporal fine structure are encoded at the output: we have described a principled way of combining such information in acoustic signal detection tasks. Furthermore, double reassignment provides a means by which the source models available in machine hearing systems can be used to obtain improved representations of target sounds in noisy acoustic environments.

APPENDIX: PSEUDOCODE

The following functions compute the probability density associated with a zero crossing interval, i , or a zero crossing interval and the peak-squared amplitude, a , jointly. Auto-covariance and autocorrelation functions, γ and ρ , are interpreted as functional arguments. $\partial(f)$ returns the functional derivative of f . The notation $\Lambda x.[f(x)]$ constructs a function f with argument x .

function INTPDF(ρ, i)

$$\rho' \leftarrow \partial(\rho)$$

$$\rho'' \leftarrow \partial(\rho')$$

$$p_I \leftarrow \frac{\rho''(i)[\rho(i)^2 - 1] - \rho(i)[\rho'(i)]^2}{2\sqrt{\rho''(0)[\rho^2(i) - 1]^3}}$$

return p_I

end function

function INTPEAKPDF(γ, i, a)

$$\rho \leftarrow \Lambda x.[\gamma(x)/\gamma(0)]$$

$$\beta \leftarrow \frac{\gamma(i) - 2\gamma(\frac{i}{2})\rho(\frac{i}{2}) + \gamma(0)}{1 + \rho(i)}$$

$$p_{A|I} \leftarrow \sqrt{\frac{a}{2\pi\beta^3}} \exp\left(\frac{a}{-2\beta}\right)$$

$$p_I \leftarrow \text{INTPDF}(\rho, i)$$

return $p_{A|I} \times p_I$

end function

If an interval/squared-peak pair (i, a) is received, the function $\text{INTPEAKPDF}(\gamma, i, a)$ will return the likelihood that it was generated by the Gaussian noise process with autocovariance function γ . By comparing the likelihoods, a detector can be constructed.

In order to detect sinusoids in noise, additional steps must be taken. The following function pre-constructs a table of fifty weights that are used in the computation of the probability of an interval given a sine-in-noise model. Here, Q is a matrix, and z and η are vectors. The function $\text{PSEUDOINVERSE}(Q, z)$ computes the pseudoinverse of Q (e.g., using Matlab's `pinv`) and multiplies it by the vector z , i.e., Q^+z .

```

function RAYLEIGHWEIGHT( $k$ )
  for  $m = 0$  to 10000 do
     $x \leftarrow \frac{m}{1000}$ 
     $z_{m+1} \leftarrow \frac{1}{\sqrt{0.02\pi}} \exp \left[ -\frac{(x-1)^2}{0.02} \right]$ 
    for  $n = 0$  to 49 do
       $\sigma \leftarrow 10^{\frac{23}{490}n-2}$ 
       $Q_{m+1,n+1} \leftarrow \frac{x}{\sigma^2} \exp \left( -\frac{x^2}{2\sigma^2} \right)$ 
    end for
  end for
   $\eta \leftarrow \text{PSEUDOINVERSE}(Q, z)$ 
  return  $\eta_k$ 
end function

```

The probability that the interval/squared-peak pair is generated by the Gaussian process with autocovariance γ plus a sinusoid with amplitude A and radial frequency ω is then computed as follows.

```

function INTPEAKSINEPDF( $\gamma, A, \omega, i, a$ )
   $p_{AI} \leftarrow 0$ 
  for  $n = 0$  to 49 do
     $\sigma \leftarrow 10^{\frac{23}{490}n-2}$ 
     $\eta \leftarrow \text{RAYLEIGHWEIGHT}(n + 1)$ 
     $\gamma_s \leftarrow \Lambda\tau \cdot [\gamma(\tau) + A^2\sigma^2 \cos(\omega\tau)]$ 
     $p_{AI} \leftarrow p_{AI} + \eta \times \text{INTPEAKPDF}(\gamma_s, i, a)$ 
  end for
  return  $p_{AI}$ 
end function

```

Suppose an interval, i , is measured from a Gaussian process with a power spectral density of Gaussian profile and 10 Hz 3 dB-bandwidth. A discriminator that decides optimally whether the process was centered at 100 Hz or 110 Hz would be coded as follows:

```

function TESTINTERVAL( $i$ )
   $b \leftarrow \underline{10}$ 
   $a \leftarrow \frac{(\pi b)^2}{4 \ln 2}$ 
   $\rho_1 \leftarrow \Lambda\tau \cdot [\exp(-a\tau^2) \cos(2\pi \cdot \underline{100}\tau)]$ 
   $\rho_2 \leftarrow \Lambda\tau \cdot [\exp(-a\tau^2) \cos(2\pi \cdot \underline{110}\tau)]$ 
   $pH_1 = \text{INTPDF}(\rho_1, i)$ 
   $pH_2 = \text{INTPDF}(\rho_2, i)$ 
  if  $pH_1 \geq pH_2$  then
    return "100 Hz"
  else
    return "110 Hz"
  end if
end function

```

This program outputs the centre frequency of the process most likely to generate the measurement (or 100 Hz in the event of a tie on evidence). A power detector is unable to make this discrimination without additional filtering. The detectors described in this paper are variations on this theme, substituting the function `INTPDF()` for the other functions above.

Matlab code that implements the above functions and reproduces all figures in the paper can be downloaded from

<https://staffwww.dcs.shef.ac.uk/people/G.Brown/interval/>

ACKNOWLEDGMENTS

Robert Mill was supported by a studentship funded by QinetiQ.

REFERENCES

- Angelsen, B. A. (1981). “Instantaneous frequency, mean frequency, and variance of mean frequency estimators for ultrasonic blood velocity doppler signals”, *IEEE Transactions on Biomedical Engineering* **28**, 733–41.
- Bom, N. and Conoly, B. W. (1970). “Zero-crossing shift as a detection method”, *Journal of the Acoustical Society of America* **47**, 1408–1411.
- Burdic, W. S. (2003). *Underwater Acoustic Systems Analysis*, 1–489, second edition (Peninsula Publishing, Incorporated).
- Chandrasekhar, S. and Sreenivas, T. V. (2005). “Auditory motivated level-crossing approach to instantaneous frequency estimation”, *IEEE Transactions on Acoustics, Speech and Signal Processing* **53**, 1450–1462.
- Cobb, S. (1965). “The distribution of intervals between zero crossings of sine wave plus random noise and allied topics”, *IEEE Transactions on Information Theory* **11**, 220–233.
- Cooke, M. P. (1991/1993). “Modelling auditory processing and organisation”, Ph.D. thesis, University of Sheffield.

- Dajani, H. R., Wong, W., and Kunov, H. (2005). “Fine structure spectrography and its application in speech”, *Journal of the Acoustical Society of America* **117**, 3902–3918.
- David, F. N. (1953). “A note on the evaluation of the multivariate normal integral”, *Biometrika* **40**, 458–459.
- Fulop, S. A. and Fitz, K. (2006). “Algorithms for computing the time-corrected instantaneous frequency (reassigned) spectrogram, with applications”, *Journal of the Acoustical Society of America* **119**, 360–371.
- Gardner, T. J. and Magnasco, M. O. (2006). “Sparse time-frequency representations”, *Proceedings of the National Academy of Sciences* **103**, 6094–6099.
- Ghitza, O. (1988). “Temporal non-place information in the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment”, *Journal of Phonetics* **16**, 109–123.
- Grigorakis, A. (1997). “Application of detection theory to the measurement of the minimum discernable signal for a sinusoid in gaussian noise displayed on a lofargram”, Technical Report DSTO-TR-0568, DSTO Aeronautical and Maritime Research Laboratory.
- Haque, S., Togneri, R., and Zaknich, A. (2007). “A temporal auditory model with adaptation for automatic speech recognition”, in *IEEE International Conference on Acoustics, Speech and Signal Processing*, volume 4, IV–1141–IV–1144.
- Higgins, R. C. (1980). “The utilization of zero-crossing statistics for signal detection”, *Journal of the Acoustical Society of America* **67**, 1818–1820.
- Kay, S. M. and Sudhaker, R. (1986). “A zero crossing-based spectrum analyzer”, *IEEE Transactions on Acoustics Speech and Signal Processing* **34**, 96–104.
- Kedem, B. (1986). “Spectral analysis and discrimination by zero-crossings”, *Proceedings of the IEEE* **74**, 1477–1493.
- Kim, D.-S., Lee, S.-Y., and Kil, R.-M. (1999). “Auditory processing of speech signals for robust speech recognition in real-world noisy environments”, *IEEE Transactions on Speech and Audio Processing* **7**, 55–69.
- Kodera, K., Gendrin, R., and Villedary, C. D. (1978). “Analysis of time-varying signals

with small BT values”, IEEE Transactions on Acoustics, Speech and Signal Processing **26**, 64–76.

Kodera, K., Villedary, C. D., and Gendrin, R. (1976). “A new method for the numerical analysis of non-stationary signals”, Physics of The Earth and Planetary Interiors **12**, 142–150.

Kumaresan, R. and Wang, Y. (2001). “On representing signals using only timing information”, Journal of the Acoustical Society of America **110**, 2421–2439.

Longuet-Higgins, M. S. (1983). “On the joint distribution of wave periods and amplitudes in a random wave field”, Philosophical Transactions of the Royal Society of London, Series A **389**, 241–258.

McAuley, R. J. and Quatieri, T. F. (1986). “Speech analysis/synthesis based on a sinusoidal representation”, IEEE Trans. Acoust. Speech Signal Process. **34**, 744–754.

McDonough, R. N. and Whalen, A. D. (1995). *Detection of signals in noise*, 1–512, second edition (Academic Press, New York; London).

Pickles, J. O. (2012). *An Introduction to the Physiology of Hearing*, 1–400, fourth edition (Emerald Group Publishing Ltd.).

Rainal, A. J. (1966). “Zero-crossing principle for detecting narrow-band signals”, IEEE Transactions on Instrumentation and Measurement **IM-15**, 38–43.

Rainal, A. J. (1967). “Another zero-crossing principle for detecting narrow-band signals”, IEEE Transactions on Instrumentation and Measurement **IM-16**, 134–138.

Rice, S. O. (1944). “Mathematical analysis of random noise”, Bell Systems Technical Journal **23**, 282–332.

Seneff, S. (1988). “A joint synchrony/mean-rate model of auditory speech processing”, Journal of Phonetics **16**, 55–76.

Sumner, C. J., O’Mard, L. P., Lopez-Poveda, E. A., and Meddis, R. (2003). “A nonlinear filter-bank model of the guinea-pig cochlear nerve: rate responses”, J Acoust Soc Am **113**, 3264–3274.

Wang, D. and Brown, G. J., eds. (2006). *Computational auditory scene analysis: Principles*,

algorithms, and applications, 1–44 (Wiley/IEEE Press).

LIST OF FIGURES

FIG. 1 Analytical and empirical distributions for the zero crossing intervals and instantaneous frequency samples of a Gaussian random process formed by passing white noise through a linear filter. A) Squared magnitude response of the filter used to generate the random process (see text). B) Autocorrelation coefficient of the resultant process. C) Cumulative distribution function of a single zero crossing interval. D) Probability density function of a single zero crossing interval. E) Probability density function of an instantaneous frequency measurement taken from two samples separated by $\Delta t = 2^{-15}$ s. In panels C–E, the analytical approximations are shown as black curves; the empirical distributions, sampled from a 300 s random signal, are shown as a grey areas. The dotted vertical lines indicate frequencies or intervals ± 0.5 octave either side of the centre frequency of the process (1 kHz). 11

FIG. 2 Joint statistics of intervals and peaks. A) Diagram illustrating how small changes to sample values around zero, $\Delta\xi_1$ and $\Delta\xi_2$, relate to small changes in crossing interval, Δi_1 and Δi_2 , provided that one assumes a functional form for the local waveform. B) Five contours of the joint interval-peak p.d.f., $p_{IA}(i, a)$, uniformly spaced between zero and its maximum. The thick grey lines are contours of the empirical distribution; the thinner black lines are those of the analytical distribution. Marginals for $p_I(i)$ and $p_A(a)$ are shown to the bottom and right, respectively. The p.d.f. is that of the example Gaussian process used in Figure 1. 16

FIG. 3 A) The distribution of the zero crossing intervals of a sinusoid in noise, based on a 300 s sample (grey area). The sinusoid has frequency 950 Hz and amplitude $\alpha \approx 127$, and the noise is white with unit power per 1 Hz band, giving a narrow-band SNR of ≈ 33 dB. The ZC intervals were measured after the mixture was passed through a filter with the magnitude response plotted in Figure 1A (right). 17

FIG. 4 Performance of the squared-envelope (thick, grey line), interval (dashed line; crosses), instantaneous frequency (dotted) and joint interval-peak (solid line; round marker) detectors, when following a minimum error criterion. The analytical and simulations results are plotted using lines and markers, respectively. No empirical performance is measured for the envelope or IF detectors. A) Detection performance as a function of SNR for two signal frequencies (left: on-centre; right: off-centre). B) Detection performance as a function of frequency for two SNRs (left: lower SNR; right: higher SNR). 21

FIG. 5 ROC curves for squared-envelope (thick, grey), interval (dashed) and joint interval-peak (solid) detectors. The analysis band is Gaussian in shape and centred at 1 kHz in all instances. The family of curves marked ‘A’ show the detection performance for a target signal of frequency 1050 Hz. The family of curves marked ‘B’ show detection performance for a target signal of frequency 1100 Hz. In both cases the NB-SNR is 30 dB. Chance performance is marked with a dotted line. 23

FIG. 6 Pseudospectra based on timing measurements for an example realisation of a Gaussian random process. A) Power spectral density of the process (upper) and squared magnitude frequency responses of the analysis filters (lower). B) Squared magnitude of output samples at each filter. C) Zero crossing interval histogram. D) Instantaneous frequency histogram. E) Zero crossings with peak amplitudes. In panels B–E, the black curves show analytical approximations of the expected profile, and the grey regions show a measurement taken from a random sample of 100 seconds. 27

FIG. 7 Example of the doubly-reassigned ZCIH. A) Power spectral density of the noise process (top) and distribution of the analysis filters (bottom). B) ZCIH (top) and DR-ZCIH (bottom). The white line shows the frequency track of a component with amplitude 40. C) Time-averaged ZCIH (grey line) and DR-ZCIH (black line) for the regions 0–3 seconds (left) and 7–10 seconds (right). The dotted line marks the stationary signal frequency. 30

FIG. 8 Frequency reassignment maps used at the three filters marked (i), (ii) and (iii) in Figure 7A. The abscissa shows the measured frequency ($\frac{1}{2t}$) and the ordinate shows the adjusted frequency. Pixel values show the probability density of a frequency adjustment (higher values are darker); vertical pixel columns show posterior distributions. Dotted black/white lines show filter centre frequencies. White diagonal lines plot a naïve (identity) remapping. The S-shaped white lines plot the distribution mean in each column (for visual purposes only, this mean is not used in the DR-ZCIH). 32