

## Research Article

# Masking of Time-Frequency Patterns in Applications of Passive Underwater Target Detection

Jüri Sildam

*Defense Research and Development Canada Atlantic, Dartmouth, NS, Canada B2Y 3Z7*

Correspondence should be addressed to Jüri Sildam, jsildam@gmail.com

Received 6 July 2009; Accepted 11 January 2010

Academic Editor: Gary Wood

Copyright © 2010 Jüri Sildam. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Spectrogram analysis of acoustical sounds for underwater target classification is utilized when loud nonstationary interference sources overlap with a signal of interest in time but can be separated in time-frequency (TF) domain. We propose a signal masking method which in a TF plane combines local statistical and morphological features of the signal of interest. A dissimilarity measure  $D$  of adjacent TF cells is used for local estimation of entropy  $H$ , followed by estimation of  $\Delta H = H_{tc} - H_{fc}$  entropy difference, where  $H_{fc}$  is calculated along the time axis at a mean frequency  $f_c$  and  $H_{tc}$  is calculated along the frequency axis at a mean time  $t_c$  of the TF window, respectively. Due to a limited number of points used in  $\Delta H$  estimation, the number of possible  $\Delta H$  values, which define a primary mask, is also limited. A secondary mask is defined using morphological operators applied to, for example,  $H$  and  $\Delta H$ . We demonstrate how primary and secondary masks can be used for signal detection and discrimination, respectively. We also show that the proposed approach can be generalized within the framework of Genetic Programming.

## 1. Introduction

In navy applications one needs to intercept and classify pings of an unknown sonar in presence of unknown number and type of interferers. Analyzed signals depending on duration in time ( $T$ ) and extent in frequency ( $W$ ) are divided into two broad categories: narrow-band (NB) ( $WT \ll 1$ ) and broad-band (BB) ( $WT \gg 1$ ). Passive detection of signals with NB and BB components is usually based on separate approaches (see, e.g., [1]). Typical waveforms in active sonar surveillance are continuous waveform at a constant frequency and amplitude, and frequency modulated waveform. These waveforms are also relevant in passive acoustical analysis, for example, in ping interception and classification. Appearance of these signals in spectrograms depends on the temporal extent of data which is used to estimate a single periodogram, on the amount of overlapping periodograms, and on the weighting window.

After defining the time-frequency (TF) parameters required for spectrogram processing, and general expectation about expected signals, one has to answer the question: what features of detected NB and BB signals are sufficient for a target classification? Extended feature analysis is required

for analysis of active echoes to discriminate between targets and clutter. A sound timbre is an example of a perception-based feature (see, e.g., [2]). However, in the case of strong interfering signals, or in the case of limited number of sensors (e.g., single sensor), isolation of the full TF signature of a target may not be possible. In such a case one may try to characterize the detected energy peaks locally, that is, not just by the frequency of detected signal peak and its amplitude but also by spectral energy distribution in close proximity of the detected peak. Classification of spectral peaks is addressed in [3, 4]. Moreover, [5] performed spectrogram segmentation by deciding whether each spectrogram point (TF coefficient) belongs to some deterministic component region or to noise (or background) region. They pointed out that to perform such a decision one needs more than the energy level. The authors chose to utilize local statistical features so that these features aggregated in feature space.

The present approach also concentrates on local TF cell statistics of an input spectrogram image  $P$ . We are interested in testing for presence of a signal of interest (SI) in a given TF cell regardless of a number and type of waveforms of interfering sources (ISs).

We use two basic assumptions: (1) SI and IS most of the time do not overlap in TF space; (2) both SI and IS exhibit a high SNR.

Typically in active radar/sonar applications a non-parametric approach is used when one hopes to create a simple detector which is insensitive to environment at a cost of some deterioration in performance as compared to the detectors optimal in Neyman-Pearson sense [6–8]. Since in our case the number and type of waveforms of IS may change, we approach this problem in the framework of machine learning of one class classification [9, 10]. However, as opposed to the machine learning approaches (e.g., support vector machine or neural networks which put out complex classifiers and are prone to over-fitting), we propose a simple two-step approach. It combines statistical processing and morphological operations to obtain time-frequency masks of SI. We show that the simplicity of our approach lies in the fact that even after visual inspection of images of proposed features, it is possible to construct masks which can be used for masking and discrimination of SI. Moreover, we show that the proposed interactive feature selection and construction of masking rules can easily be generalized in a framework of Genetic Programming (GP) [11].

The proposed approach will be described in Section 2. Section 3 will present examples of the proposed approach. While the first three examples will show details of the approach where feature manipulation is based on visual image inspection, the last example will present performance measures over a wide range of peak-to-peak ratios based on a mask created using GP. Finally, a summary and conclusions will be given.

## 2. Approach

**2.1. Overview.** A block-scheme of the proposed approach is shown in Figure 1. We start the explanation of the proposed approach with a short summary followed by a detailed explanation of each step.

The approach consists basically from two steps: the statistical image processing of one or more spectrograms  $P$  (Figure 1A) and the construction of masks that comprise feature selection and their combination using some heuristic rules (Figures 1B and 1C). More than one spectrogram may be needed for inspection of feature images under changing environment. At the end of this section we show that a combination of features and heuristic rules can be presented in a form of a syntax tree of GP, which provides a framework for generalization of SI masking.

The first part of statistical processing is based on the Maximum Mean Discrepancy test (MMD) which is carried out on a pair of TF cells within a sliding TF window. By moving the TF window at maximum time and frequency resolution, correlated MMD estimates of dissimilarity  $D$  of TF cells are obtained. To estimate the information content of  $D$  distribution (Figure 1A), the local entropies  $H$  and  $\Delta H$  which differ in their TF support, respectively, are calculated. As a result the six images (the reason for six instead of three images will be apparent from the next section) of the features

of estimated spectrogram are used to define a set of feature vectors  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$  (Figure 1B).

**2.2. Maximum Mean Discrepancy.** We apply the Maximum Mean Discrepancy test (henceforth *dissimilarity*) on a pair of adjacent time-frequency cells in a TF window with a predefined time and frequency support. We apply this test to quantify dissimilarity of neighboring TF cells. Theoretical background and details about the MMD test can be found in [12]. An empirical biased estimate of MMD defined for the pair of TF cells, centered at  $t_1, f_1$  and  $t_2, f_2$ , respectively, can be written as

$$D[C(t_1, f_1), C(t_2, f_2)] = \left[ \frac{1}{m^2} \sum_{i,j}^{m,n} k(\mathbf{x}_i, \mathbf{x}_j) - \frac{2}{mn} \sum_{i,j}^{m,n} k(\mathbf{x}_i, \mathbf{y}_j) + \frac{1}{n^2} \sum_{i,j}^{m,n} k(\mathbf{y}_i, \mathbf{y}_j) \right], \quad (1)$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are vectors defined on normalized values of power spectral density  $\hat{c}(t_k, f_i)$  of the TF window of the first and second TF cells, respectively, (see Figure 2), and  $k(\mathbf{x}_i, \mathbf{y}_j)$  is a kernel function. In this work we used the Gaussian radial basis function  $k(\mathbf{x}_i, \mathbf{y}_j) = \exp(-\|\mathbf{x}_i - \mathbf{y}_j\|^2/\sigma^2)$ , where  $\sigma$  is a scaling parameter. Values of the dissimilarity measure (1) vary between 0 and 1. The TFW intensity values  $c(t_i, f_j)$  were normalized by the  $L_2$  norm, that is,

$$\hat{c}(t_k, f_i) = \frac{c(t_k, f_i)}{\sqrt{\sum_{i=0}^{M'-1} \sum_{j=0}^{N'-1} \|c(t_i, f_j)\|^2}}. \quad (2)$$

The MMD test of a pair of TF cells can be carried out along frequency or time axes. In the former case, both cells have a common time span, with a mean time  $t_c$  and different frequency intervals. In the latter case, both cells have a common frequency span, with a mean frequency  $f_c$ , and different time intervals (see Figure 2). For BB signals the highest gradients of spectral power are observed along time axis. Conversely for the NB tonals and associated harmonics, the highest gradients are observed along the frequency axis. We denote the corresponding MMD tests broad-band ( $D_{BB}$ ) and narrow-band ( $D_{NB}$ ) detection tests, respectively. However, there will be cases when BB signals extend only over a certain portion of spectrogram, and have sharp gradients both in time and frequency, or when NB signals are not pure tonals, as in the case of frequency-modulated signals. In such cases,  $D_{NB}$  and  $D_{BB}$  tests detect also portions of BB and NB signals, respectively. The respective MMD tests are then given by the equations:

$$\begin{aligned} D_{NB}(t_c, f_c) &= D[C(t_c, f_1), C(t_c, f_2)], \\ D_{BB}(t_c, f_c) &= D[C(t_1, f_c), C(t_2, f_c)]. \end{aligned} \quad (3)$$

By moving a TF window step by step at the time and frequency resolutions of the spectrogram, one obtains both  $D_{NB}$  and  $D_{BB}$  distributions. Each of the distributions will exhibit locally correlated values in the TF plane. Inspecting numerous  $D$  (below we will drop subscripts BB or NB when

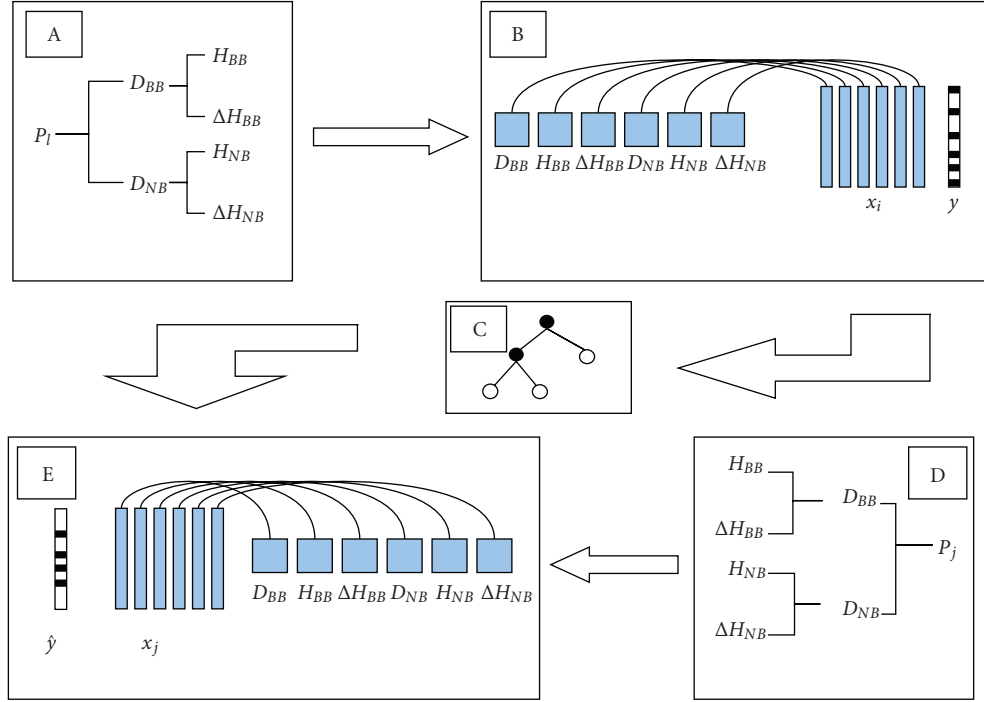


FIGURE 1: The Statistical (A, D) and Morphological (B, C, E) Data Processing Flow. A—Spectrograms  $\{P_1, \dots, P_k\}$  are a subject to MMD test, the output of which  $D_{NB}$  and  $D_{BB}$  are used for estimation of entropy  $H$  and difference of entropies  $\Delta H$ ; B—The images of features (filled squares) are used to form  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)^T$ . The feature vectors  $\mathbf{x}$  corresponding to SI are labeled by  $y = 1$  and by  $y = 0$  in the opposite case; C—mask of signal of interest in a form of GP tree; D—statistical processing of a test spectrogram  $P_j$ ; E—features estimated in D are used for testing the masking algorithm C.

talking about both NB and BB changes) distributions, it was noticed that for certain values of the Gaussian scaling index  $\sigma^2$  of the kernel,  $D$  exhibits only small changes in statistically homogeneous areas and significant changes in areas of strong spectral energy variations. Usually in the MMD test the scale factor  $\sigma^2$  is tuned so that no change would correspond to 0 and a significant change would correspond to 1. However, since reusable  $\sigma^2$  values are desired, rather than fine tuning the scale parameter,  $D$  distributions were inspected based on a constant  $\sigma^2$  using a diverse dataset. For  $\sigma^2 = 0.0001$ , the areas of relatively homogeneous distribution of power spectral density were close to some constant value (e.g., 0.4), while in the areas of significant changes  $D$  values fluctuated relative to this average. Entropy ( $H$ ) estimation provides a consistent way to quantify information related to these changes, irrespectively of actual  $D$  values. As in case of  $D$  estimation, the estimated  $H$  had also a local TF support (an example will be shown in Section 3).

**2.3. Entropy Estimation.** Since the changes observed in a spectrogram were divided into two broad categories of NB and BB signals, entropy distributions for both NB and BB tests were estimated. Besides  $H$  estimation, based on the local 2D TF support, 1D support of entropy ( $\hat{H}$ ) was performed. A cartoon summarizing time-frequency support of all features

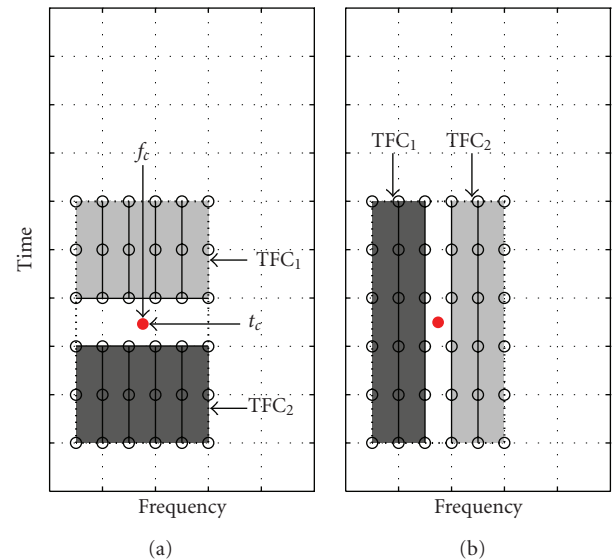


FIGURE 2: Cartoon of a pair of time-frequency cells ( $TFC_1$  and  $TFC_2$ ) placed in a time-frequency window (surrounded by the thick dotted line) is shown using two different gray shades. Panels (a) and (b) show relative location of TFCs used for dissimilarity estimation of BB and NB changes, respectively. The red filled circles in the middle of the TF cells correspond to the mean values of time ( $t_c$ ) and frequency  $f_c$  of respective TF window.

Variable	TF support	Variable	TF support
$D_{BB}$		$D_{NB}$	
$H_{BB}$		$H_{NB}$	
$\Delta H_{BB}$		$\Delta H_{NB}$	

FIGURE 3: Cartoon summarizing TF support of statistical features.

is shown in Figure 3. To compare directional  $\hat{H}$  changes, the  $\hat{H}$  estimation was carried out pairwise; entropies were estimated along a time axis at a mean frequency and along a frequency axis at a mean time so that the respective pair had a common reference point centered in the TF window. Moreover, by limiting the number of grid points used for entropy estimation along a grid line in a single direction, the number of entropy values was also limited. In this work, the smallest number of points, three (one to each side of a reference point), was used, that is,

$$\begin{aligned} \hat{H}_{NB}(k, l) &= -\sum_j \hat{p}(k, j) \log p(k, j), \\ \text{where } j &= \{l-n, l, l+n\}, \\ \hat{H}_{BB}(k, l) &= -\sum_k \hat{p}(i, l) \log p(i, l), \\ \text{where } i &= \{k-m, k, k+m\}, \end{aligned} \quad (4)$$

where the probabilities  $\hat{p}(k, j)$  were based on histograms estimated using three points (e.g.,  $l-n$ ,  $l$ , and  $l+n$ ):

$$\begin{aligned} \hat{p}_{k,j} &= \frac{\hat{W}_{k,j}}{N}, \quad \text{where } j = \{l-n, l, l+n\}, \\ \hat{p}_{i,l} &= \frac{\hat{W}_{i,l}}{N}, \quad \text{where } i = \{k-m, k, k+m\}. \end{aligned} \quad (5)$$

The sign of the difference between the pair of entropies, calculated in time and frequency directions, stresses either BB or NB changes, respectively:

$$\begin{aligned} \Delta \hat{H}_{NB} &= \begin{cases} H_{NB} - H_{BB}, & \text{if } H_{NB} - H_{BB} > 0, \\ 0 & \text{otherwise,} \end{cases} \\ \Delta \hat{H}_{BB} &= \begin{cases} H_{BB} - H_{NB}, & \text{if } H_{BB} - H_{NB} > 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned} \quad (6)$$

Note that the constraint  $\Delta \hat{H} = 0$  is used here only for simplicity; later within the GP framework it will be removed. By using all values with the same sign and including 0, the number of possible values of  $\Delta \hat{H}$  is four. In the next section it will be shown how the respective sets can be used to obtain binary masks. These masks are the primary binary masks since these masks are sufficient for signal detection purposes. In presence of strong interferences (e.g., jamming [13, 14]), discrimination of detected signals is required. This motivated

creation of a second set of binary masks. Again, derivation of the secondary masks should be local in the TF plane with a possibility of global optimization of the features, used in derivation of the binary masks, for global discrimination analysis. One of such features, the local entropy based on two-dimensional  $D$  distribution, was calculated, that is,

$$\begin{aligned} H(k, l) &= -\sum_{i,j} p(i, j) \log p(i, j), \quad \text{where } i = 1, \dots, N; j = 1, \dots, M, \end{aligned} \quad (7)$$

where probability  $p(i, j)$  was defined using one-dimensional histogram  $W_{k,l}$  based on stacked data of two-dimensional  $D$  distribution values within a TF window:

$$p_{k,l} = \frac{W_{k,l}}{NM}. \quad (8)$$

#### 2.4. Construction of Time-Frequency Masks

**2.4.1. Visual Analysis.** Taking  $N = 3$  in (5), the primary masks were constructed using the following set  $\Delta \hat{H} = \{0, 0.4917, 0.4597, 0.9514\}$ . Below we will use three primary masks:

$$\begin{aligned} \Delta \hat{H}_1 &= \{0\}, \\ \Delta \hat{H}_2 &= \{0.4917, 0.4597\}, \\ \Delta \hat{H}_3 &= \{0.9514\}. \end{aligned} \quad (9)$$

To construct the secondary masks four generic rules were used. For the feature matrices  $A(k, l)$  (e.g., NB entropy),  $B(k, l)$  (e.g., BB entropy), and  $C(k, l)$  (e.g., spectral data) these generic rules were as follows:

$$\begin{aligned} F_1[A(k, l) \mid n] &= \{(A(k, l-n) = A(k, l+n)) \vee (A(k, l-n) > 0)\}, \\ F_2[B(k, l), C(k, l) \mid n, m] &= \{(B(k-m, l) < B(k, l)) \vee (B(k+m, l) < B(k, l)) \vee (C(k, l) > 0)\}, \\ F_3[A(k, l), B(k, l), C(k, l) \mid n, m] &= F_1[A \mid n] \vee F_2[B, C \mid n, m], \end{aligned} \quad (10)$$

where  $k = 1, \dots, K$  and  $l = 1, \dots, L$  refer to the time and frequency indices, respectively. The fourth rule was based on (primary) masked data with the increasing span of the structuring element:

$$\begin{aligned} F_4[A(k, l), B(k, l) \mid n] &= \max[A(k, l-n), \dots, A(k, l), \dots, A(k, l+n)] \vee B(k, l) > 0. \end{aligned} \quad (11)$$

In the next subsection we will point out a way to generalize the construction of masks given above.

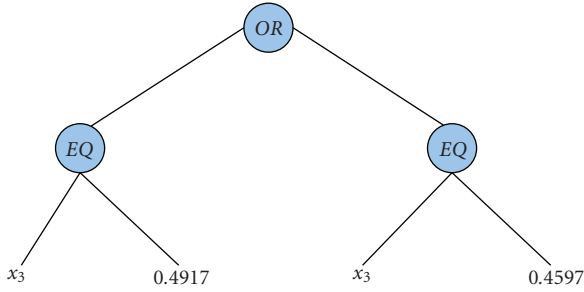


FIGURE 4: The primary mask  $\Delta\hat{H}$  (9) presented in a form of the syntax tree.

**2.4.2. Genetic Programming.** In Genetic Programming programs are represented as syntax trees. A tree includes nodes, which indicate instructions to execute, and links which point to inputs (or terminals) used for execution of instructions [11]. Initially a random set (or population) of trees (or programs) is created and executed recursively following chosen instructions. Using some fitness measure on results of program execution, only a few programs are selected which are then used to generate new programs using some predefined genetic operations. Selection of the fittest programs followed by genetic generation of new programs is repeated until some predefined criteria are met. The final result is given in a form of syntax tree.

In our case, any (i.e., primary and secondary) of the masks developed above can be presented as a syntax tree. For simplicity we take a primary mask  $\Delta\hat{H}$  defined in the previous section. The feature images (both NB and BB)  $d_{i,j} \in D$ ,  $h_{i,j} \in H$ , and  $\hat{h}_{i,j} \in \hat{H}$  can be used to form feature vectors  $\mathbf{x}_{i,j} = (\mathbf{d}_{\text{NB}}, \mathbf{h}_{\text{NB}}, \mathbf{\hat{h}}_{\text{NB}}, \mathbf{d}_{\text{BB}}, \mathbf{h}_{\text{BB}}, \mathbf{\hat{h}}_{\text{BB}})_{i,j} = (x_1, \dots, x_6)_{i,j}$  (note that in most practical cases a feature vector will be defined not just at a single pixel  $i, j$  but also at neighboring  $i-n, i+n, j-m, j+m$  pixels). The members of  $x$  can be then used as inputs to the syntax tree (Figures 1B and 1E), which in turn can be evaluated iteratively. An example of primary mask presented in a form of the syntax tree is shown in Figure 4. In the present case, selection of nodes, inputs, and links forming the tree is carried out interactively. It is straightforward in a case of construction of primary masks (9) and is more complicated for the secondary masks. However, by using the masking approach in GP framework, a generalized method for detection and discrimination of IS can be obtained. In the next section we will show three concrete examples based on combination of primary and secondary masks developed earlier. The proposed approach will be then generalized in GP framework to show its ability to work in a wide range SI-IS peak-to-peak ratios.

### 3. Algorithm Development and Application

**3.1. Construction of Masks Using Visual Analysis.** All given examples consider the case when only a single sensor is available for recording of acoustical data. We construct two masks: primary and secondary. By applying the primary

mask, we expect to detect the signals of interest along with the interfering signals. The secondary morphological mask is used then to discriminate between a deterministic SI, random NB Gaussian, and BB interfering signals. Three cases will be considered. The first case simulates a deterministic damped sinusoidal signal against a background of weaker random Gaussian interference and recurring BB signals (e.g., modeling pings of own ship echo-sounder). In the second, a more difficult case, the situation is reverse. Now the NB interference is stronger than the recurring signal of interest at the background of uniform noise and recurring BB signal. Finally, a recording of a bird song [15] will be considered. We chose the bird song instead of, for example, marine mammal because the corresponding file is accessible from the Internet, so that the interested parties can carry out similar calculations using the same data. The bird song example also exhibited a mixture of NB and BB features, from which we were able to extract NB components.

The raw time series were simulated as a sum of (i) NB Gaussian pulses (NB interference) at randomized central frequencies, which were generated by passing white noise through a narrow-band digital filter [16]; (ii) broadband Ricker wavelets (BB interference); (iii) deterministic damped CW pulses (signal of interest) recurring at the time interval 0.53 s and central frequency 7500 Hz; and (iv) uniform white noise. Two different examples were generated using different amplitude ratios of the components. Values of the parameters used in the estimation of spectrogram features are given in Table 1.

The ratios of the peak values of signals, as well as signal versus background, are shown in Table 2. Figures 5(a) and 5(b) show the respective spectrograms. Note that while the signal and background noise ratios are different in Figures 5(a) and 5(b), the geometrical appearance of the signal as compared to the interfering random signal has not been changed. In both cases the signal bandwidth (defined as 15 dB decrease from the signal's peak value) is about 800 Hz. A similar decrease in the interfering signal is observed at a bandwidth equal to roughly a doubling in the value of the spectral resolution (i.e., 117 Hz). Also note that while the signal duration is comparable with the spectrogram time resolution (one pixel), the duration of the NB interfering signal is about 4 pixels (i.e.,  $0.0171 \cdot 4 = 0.0684$  s). Finally, in both cases BB interference could be observed over most of the frequency span of spectrogram with a duration of 0.0171 second (1 pixel).

An example of the distribution of features used in present work is shown in Figure 6 for detections of strong (a) and weak (b) signals. An example of local TF support used in the first experiment is shown by the red boxes ( $D$  and  $H$  support) and red crosses ( $\Delta\hat{H}$  support). Information about the geometrical appearance of underlying distributions was used to construct secondary binary masks based on local statistical and morphological operators. Below it will be shown that while the primary mask can be used effectively for detection of signal of interest as well as interfering signals, the secondary masks can be used for reduction of false alarms due to interfering signals. Detection of NB or BB signals



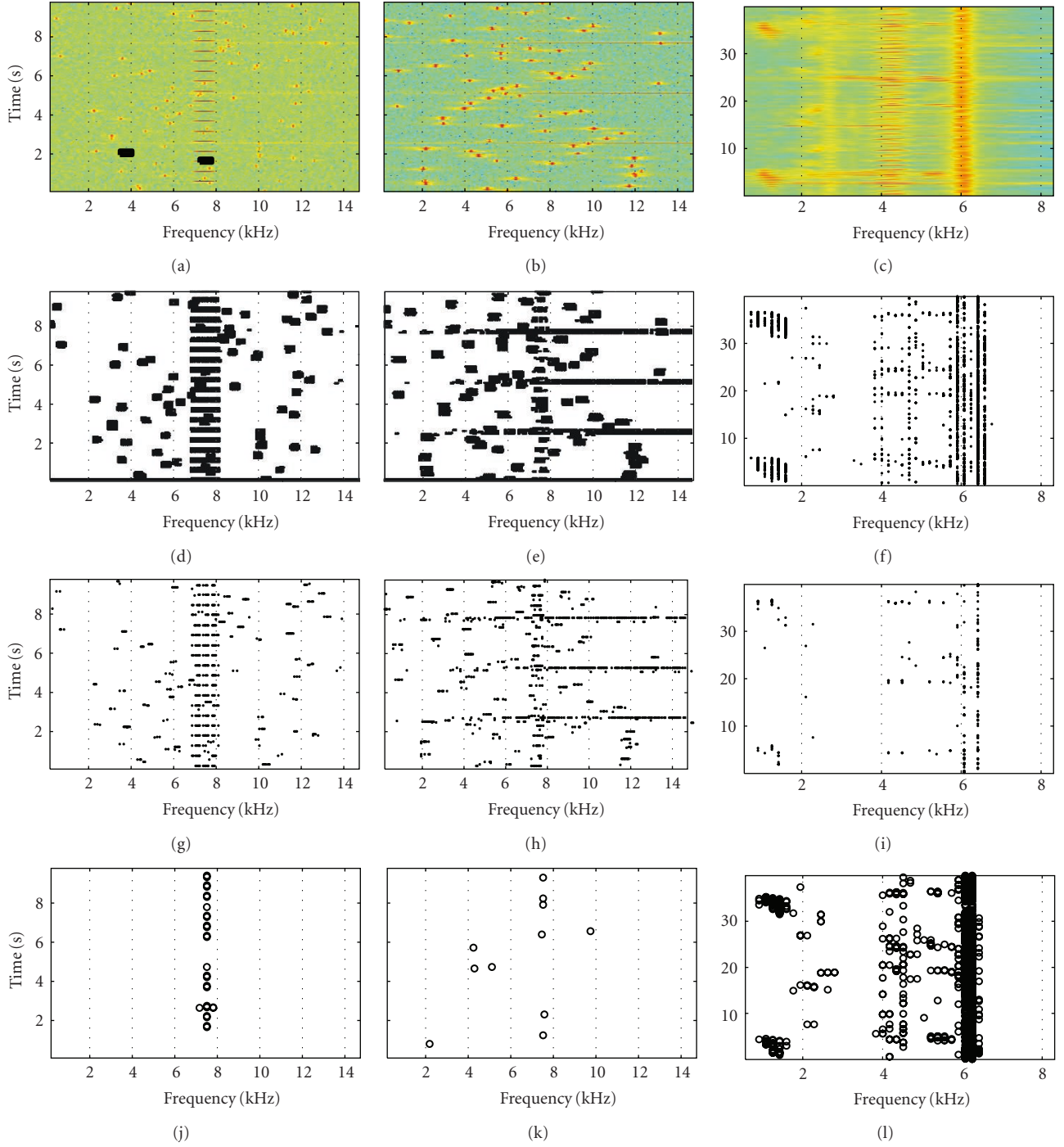


FIGURE 5: The rows from top to bottom: spectrogram, the first primary mask, the second primary mask, and the secondary mask for experiments given by columns from left to right: one (weak NB interfering signal), two (strong NB interfering signal), and three (bird call).

TABLE 1: Meta-parameters used in estimation of spectrogram features:  $D_r$ ,  $D_c$ ,  $H_r$ ,  $H_c$ ,  $\Delta H_r$ ,  $\Delta H_c$  are number of pixels used to define TF windows of  $D$ ,  $H$ , and  $\Delta H$ , respectively;  $\Delta F$  and  $\Delta T$  show spectrogram frequency and time resolution;  $F_s$ —sampling frequency.

Experiment Name	$D_r$	$D_c$	$H_r$	$H_c$	$\Delta H_r$	$\Delta H_c$	$F_s$ (Hz)	$\Delta F$ (Hz)	$\Delta T$ (s)
First	6	10	6	3	1	1	30000	58.6	0.0171
Second	6	10	6	3	1	1	30000	58.6	0.0171
Third	6	6	6	5	3	1	44100	172.3	0.0058

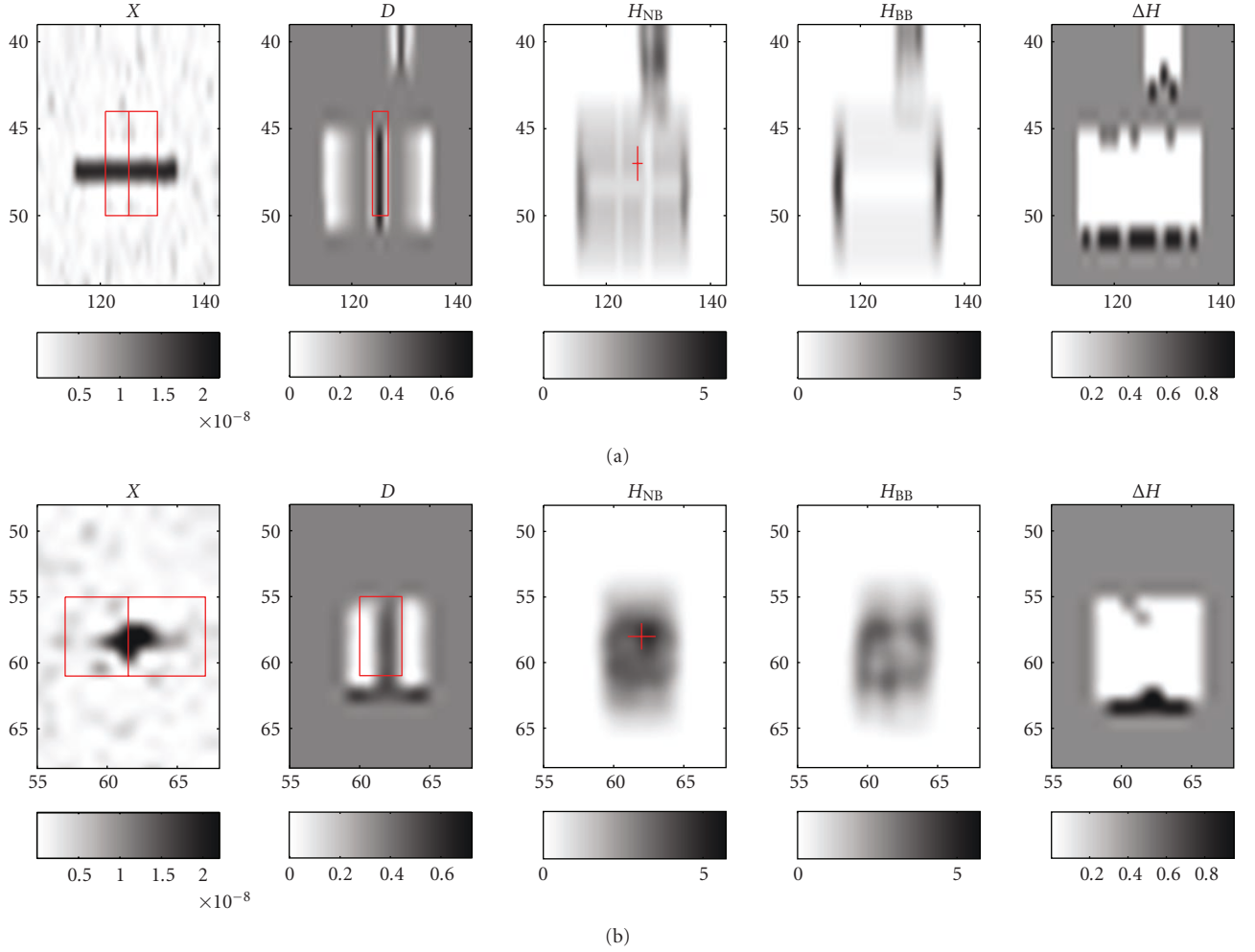


FIGURE 6: Local distribution power spectral density ( $X$ ), dissimilarity  $D_{NB}$ , entropies  $H_{NB}$  and  $H_{BB}$  and entropy difference  $\Delta\hat{H}$  for the areas marked by black boxes in Figure 5(a). The upper row corresponds to the box covering a signal of interest at 7500 Hz, and the lower row corresponds to the interfering signal covered by the lower frequency. The two red rectangles and the red cross in the figure correspond to the TF windows needed for  $D$  test (leftmost panels), and  $H$  estimation (the second column from the left), respectively. Note that scaling in pixels is different for the upper and lower rows.

TABLE 2: Values of Deterministic Signal-to-background noise (SI/BN), random NB interference to BN (RI/BN), deterministic broadband interference to BN (BI/BN), and DS to random NB interference (SI/RI) for experiments First and Second.

Experiment name	DS/BR (dB)	RS/BR (dB)	BI/BR (dB)	DS/RI (dB)
First	46	30	11	16
Second	20	40	24	-20

in the context of the present approach has the meaning of carrying out a dissimilarity test in frequency or in time direction, respectively. From point of view of these tests, the BB signals with  $WT > 1$  but  $WT \ll FsT/2$  can be regarded as either NB or BB signals depending on a final goal. Although both  $D_{NB}$  and  $D_{BB}$  tests were performed in each experiment, the primary mask was based on only one. In the

first two examples  $D_{BB}$  was used, and in the third test with a bird call,  $D_{NB}$  was used.

In all three experiments the NB detection given by (3) was used to derive the primary masks. The results of the application of the primary masks are shown in the two middle rows of Figures 5(d) and 5(e) were obtained using the primary mask  $\Delta\hat{H}_1$ , Figure 5(f) using  $\Delta\hat{H}_2$ , and Figures 5(g)–5(i) using  $\Delta\hat{H}_3$  of (9). Figures 5(j)–5(l) were obtained using secondary masks given in the order of the experiment number (1 corresponds to (j); 2 to (k), and 3 to (l)). The respective secondary mask implementation was given by

- (1)  $F_3(H_{NB}, H_{BB}, \Delta\hat{H}_{BB} \mid n = 1, m = 1) \vee F_1(H_{BB}^T \mid k = 3)$ ,
- (2)  $F_3(H_{NB}, H_{BB}, \Delta\hat{H}_{BB} \mid n = 3, m = 1)$ ,
- (3)  $F_4(H_{NB}, \Delta\hat{H}_{NB} \mid n = 1) \cap F_4(H_{NB}, \Delta\hat{H}_{NB} \mid n = 2), \dots, \cap F_4(H_{NB}, \Delta\hat{H}_{NB} \mid n = 6)$ ,

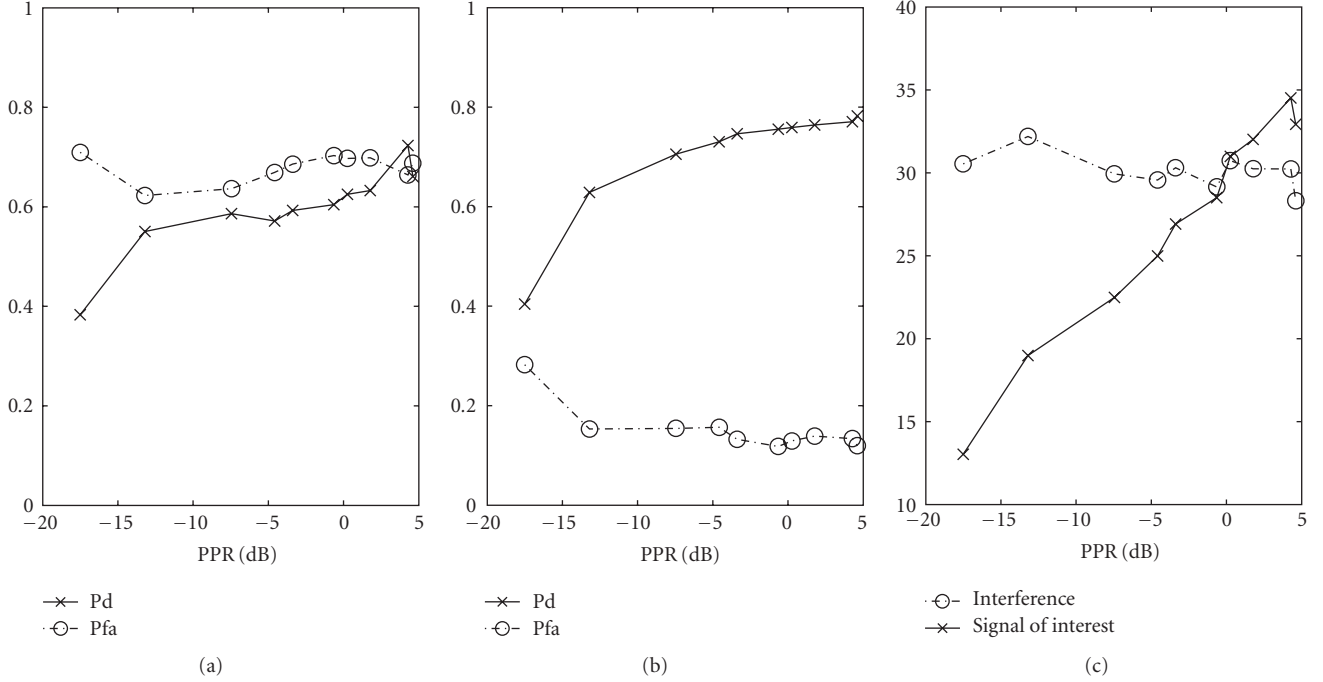


FIGURE 7: (a) results of interactive masking. Shown are the detection (Pd) and false alarm (Pfa) probabilities as a function of peak-to-peak ratio (PPR) of SI and IS; (b) the same as in A but for automated masking; (c) SNR of signal of interest and interfering signal; All results are based on 500 runs, 50 runs per each value of PPR.

where superscript  $T$  implies a matrix transpose. While in the first (strong signal) and the third (bird call) cases different combinations of secondary masking gave quite consistent results with different choices of a secondary mask, in the case of the second experiment with strong interference and weak signal the results of false alarm reduction were sensitive to the choice of the secondary mask. In general one can see that the primary masks worked well for detection of signal of interest in all three experiments.

Visual inspection shows that in the first experiment most of the NB signals were detected, while BB interference remained undetected as required (except the bottom row, shown by a continuous black line, which was related to the reduced number of grid points at the beginning of masking). However, in the second experiment the BB lines can be clearly seen in Figures 5(e) and 5(h). One should recall that since the signal of interest was on a border line between NB and BB signals, for experiments one and two we used a combination of  $D_{BB}$  and  $H_{NB}$  tests. In the third experiment however, we used a combination of  $D_{NB}$  and  $H_{NB}$  tests to derive the primary masks. In this case, already the first primary mask was useful for detection of bird call vocalization around 4200 and 6200 Hz. The secondary mask performed simple dilation of the masked data. In the first experiment, the secondary mask successfully masked interfering signals. The second experiment was the most difficult case. In addition to the secondary mask, a median filter was applied, calculated as an absolute difference between median  $H_{BB}$  and  $H_{NB}$  values estimated for each row of the spectrogram. The rows with the respective difference

higher than 0.05 were set to zero. As a result, dominant detections were observed as expected at 7500 Hz. In the next section we will show that significant reduction of false alarms can be achieved via mask construction in the framework of GP.

**3.2. Performance and Automation of Masking of Time-Frequency Patterns.** To evaluate masking performance we modified previously described simulation so that now it incorporated changing SNR of signal of interest. Each run that lasted about 10 seconds and that was carried out at constant SNR was repeated 50 times. Then the SNR was increased and another 50 runs were performed. Ten different SNR values were used. The range of SNRs values was chosen so that ratio of maximums of spectral peaks of SI and IS (PPR) changed roughly from -20 to 5 dB. Since SI was deterministic, the number of peaks over 10 second period was constant and equaled to 19. To eliminate redundancy in a number of masked peaks, the respective peaks were clustered using kd-tree closest point search method. The detection probability (Pd) was defined as a ratio between the number of detected (masked) clustered peaks and the true number of peaks, and the false alarm probability was defined as a ratio between of the number of incorrect and number of detected clustered peaks.

To automate masking of TF patterns we used an open source Matlab GP toolbox [17]. According to Koza [11], a single set-up of governing parameters of GP can be used to address a wide range of problems. The toolbox's demo script "demoparity.m" was adopted to accommodate for the format



of input features vectors and for a list of functions. After these changes, we were able successfully automate TF pattern masking.

The feature vectors centered at a pixel  $i, j$  were defined as  $\mathbf{x}_{i,j} = (\mathbf{d}_{\text{NB}}^T, \mathbf{h}_{\text{NB}}^T, \hat{\mathbf{h}}_{\text{NB}}^T, \mathbf{d}_{\text{BB}}^T, \mathbf{h}_{\text{BB}}^T, \hat{\mathbf{h}}_{\text{BB}}^T)_{i,j} = (\mathbf{x}_1, \dots, \mathbf{x}_{18})_{i,j}$ , where, for example, the first six NB members were  $x_1 = d_{i,j-1}, x_2 = d_{i,j}, x_3 = d_{i,j+1}, x_4 = h_{i,j-1}, x_5 = h_{i,j}, x_6 = h_{i,j+1}$ . The transposed feature vectors formed a matrix with a row labeled by  $y_i = 1$  or by  $y_i = 0$  depending on presence or absence of SI, respectively. The set of Matlab functions used by GP was  $\{\text{AND}, \text{OR}, \text{NOT}, \text{GE}, \text{LE}, \text{PLUS}, \text{MINUS}, \text{EQ}\}$ . The number of populations was set to 350 and program was stopped after 20 generations.

For automated creation of a TF mask we used 10 spectrograms, each corresponding to a different SNR. The respective spectrograms were processed following the proposed approach (Figure 1). To show the improvement gained by automation of TF masking of SI, we compare the respective results with the results obtained by processing of the same data set using a combination of two primary masks. The masks were constructed using a three by three pixels matrix of  $\Delta\hat{H}$  centered at a given time and frequency of SI obtained at  $-15$  and  $0$  dB PPR, respectively. Presence of SI was declared when either of the primary masks matched exactly underlying TF pattern of  $\Delta\hat{H}$ .

Comparing the results of interactive and automated masking (Figures 7(a) and 7(b)) one can see that while in both cases the detection probability was relatively high almost over the whole range of PPRs (except the first estimates at  $-18$  dB), false alarm rate was significantly lower in the case of automated masking (Figure 7(b)). The SNR of the BB signal was kept approximately constant and equal to the value used in the previous experiment.

**3.3. Qualitative Comparison of the Proposed Approach to Other Approaches.** The first part of this work separated the problem of signal detection and discrimination into two parts, respectively, by defining primary and secondary masks. Finally we showed that both of these problems could be addressed within a single framework of GP.

Separating the problem into two steps is useful in showing how preliminary information can be incorporated into the construction of feature vectors. In this work we assumed that all signals can be presented in one of the two categories: either NB or BB signals. The statistical MMD test followed by estimation of entropies capitalized on this idea by producing features  $D$ ,  $H$ , and  $\Delta H$ . Clearly the proposed use of MMD test is not the only test that can be used for signal detection in spectrograms. In cases when one has an ability to analyze recorded data over time periods sufficient for track detection, or when signal detection is a part of a tracking approach, a number of approaches can be used for track detection. A recent overview of track detection in spectrograms is given in [18]. More research is required to understand how information gained during signal or track detection can be used for signal or track discrimination. Therefore approaches developed for track detection in spectrograms (e.g., active contour models [19], reassignment, Radon

[20] and Hough [21] transforms, simple thresholding) should be complemented by statistical information (e.g., higher statistical moments) obtained in close proximity of detected time-frequency tracks. Additionally, the features based on cepstral processing may be used when one is working with quasiperiodic signals, for example, sonar pings.

For purposes of this work, postprocessing of collected data implicitly provided statistical and morphological information mainly about signal of interest. Since a number and type of interfering sources were assumed to be unknown, the presented problem can be treated as a one-class classification or novelty detection problem. Presented results show that by combining statistical and morphological approaches it is possible to reduce false alarms without the reduction of signal detection rate. It is interesting to note that recently in active sonar measurements the authors of [22] were able to reduce significantly false target detections due to clutter using the two-step statistical-morphological processing.

Finally, it should be noted that the computational complexity  $O(m+n)^2$  of MMD test (1) is the highest among all processing blocks involved in masking of the test data.

## 4. Summary and Conclusions

A new approach has been presented for binary masking of signal of interest in presence of narrow-band and broadband interferences in spectrograms. All computations of dissimilarities of TF cells and related entropies required for a binary mask estimation were local in time-frequency space. Rather than thresholding spectrogram data, or related dissimilarity or entropy distributions, the number of grid points used in estimation of entropies and their differences was limited. As a result, the number of values of the respective differences was limited, and these were used to define the primary binary masks, which did not require use of any thresholds for their estimation. An advantage of using dissimilarity rather than spectral energy distribution for entropy estimation was in the known scaling of  $D$ , which allowed predefinition of the histogram bin limits. We have shown that while local entropy differences calculated from the  $D$  distribution is an effective way for general signal detection (including interfering signals), then morphological operations can be used to reduce false alarms due to interfering signals. The division of the proposed masking approach into primary and secondary masks is useful if one has to make a distinction between the signal detection and discrimination, or during the interactive construction of signal discrimination processing flow. The presented approach is flexible and can be adopted depending on the underlying problem within a framework of Genetic Programming which unifies the proposed construction of two masks.

## Acknowledgment

A part of this work was supported by the NATO Underwater Research Centre.

## References

- [1] S. Pasupathy and P. M. Schultheiss, "Passive detection of Gaussian signals with narrow band and broad band components," *The Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 917–921, 1974.
- [2] V. W. Young and P. C. Hines, "Perception-based automatic classification of impulsive-source active sonar echoes," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 1502–1517, 2007.
- [3] M. Zivanovic, A. Röbel, and X. Rodet, "A new approach to spectral peak classification," in *Proceedings of the 12th European Signal Processing Conference (EUSIPCO '04)*, pp. 1277–1280, Vienna, Austria, 2004.
- [4] M. Zivanovic, A. Röbel, and X. Rodet, "Adaptive threshold determination for spectral peak classification," *Computer Music Journal*, vol. 32, no. 2, pp. 57–67, 2008.
- [5] C. Hory, N. Martin, and A. Chehikian, "Spectrogram segmentation by means of statistical features for non-stationary signal interpretation," *IEEE Transactions on Signal Processing*, vol. 50, no. 12, pp. 2915–2925, 2002.
- [6] J. B. Thomas, "Nonparametric detection," *Proceedings of the IEEE*, vol. 58, no. 5, pp. 623–631, 1970.
- [7] P. P. Gandhi and S. A. Kassam, "Analysis of CFAR processors in homogeneous background," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 427–445, 1988.
- [8] H. Chen, P. K. Varshney, S. Kay, and J. H. Michels, "Noise enhanced nonparametric detection," *IEEE Transactions on Information Theory*, vol. 55, no. 2, pp. 499–506, 2009.
- [9] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [10] D. Tax and R. Duin, "Uniform object generation for optimizing one-class classifiers," *Journal of Machine Learning Research*, vol. 2, pp. 155–173, 2001.
- [11] J. R. Koza, *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, The MIT Press, Cambridge, UK, 1992.
- [12] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Schölkopf, and A. Smola, "A kernel method for the two-sample problem," *Journal of Machine Learning Research*, vol. 1, pp. 1–10, 2008.
- [13] X. Mankun, P. Xijian, L. Tianyun, and X. Mantian, "A new time-frequency spectrogram analysis of FH signals by image enhancement and mathematical morphology," in *Proceedings of the 4th International Conference on Image and Graphics (ICIG '07)*, pp. 610–615, 2007.
- [14] J. Liu, J. P. Y. Lee, L. Li, Z.-Q. Luo, and K. M. Wong, "Online clustering algorithms for radar emitter classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, pp. 1185–1196, 2005.
- [15] N. Athanas, Brazil, <http://www.xeno-canto.org/recording.php?XC=11613>.
- [16] B. H. Maranda, "A method for generating narrowband test signals," Technical Memorandum DRDC-ATLANTIC-TM-2002-193, Defence R&D, Atlantic, Canada, 2002.
- [17] S. Silva, "Gplab—a genetic programming toolbox for MATLAB," <http://gplab.sourceforge.net/index.html>.
- [18] T. A. Lampert, "A survey of spectrogram track detection algorithms," *Applied Acoustics*, vol. 71, no. 2, pp. 87–100, 2010.
- [19] T. A. Lampert and S. E. M. O'Keefe, "An active contour algorithm for spectrogram track detection," *Pattern Recognition Letters*. In press.
- [20] A. C. Copeland, G. Ravichandran, and M. M. Trivedi, "Localized radon transform-based detection of ship wakes in SAR images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 33, no. 1, pp. 35–45, 1995.
- [21] T. L. Dixon and L. H. Sibul, "A parameterized hough transform approach for estimating the support of the wideband spreading function of a distributed object," *Multidimensional Systems and Signal Processing*, vol. 7, no. 1, pp. 75–86, 1996.
- [22] G. Ginolhac, J. Chanussot, and C. Hory, "Morphological and statistical approaches to improve detection in the presence of reverberation," *IEEE Journal of Oceanic Engineering*, vol. 30, no. 4, pp. 881–899, 2005.