A statistical approach to RDX detection with THz reflection spectra

Ivan Zorych,[1,*] Yew Li Hor,[2] Alexander M. Sinyukov,[2] Zoi-Heleni Michalopoulou,[1] Robert B. Barat,[3] Dale E. Gary,[2] and John F. Federici[2]

[1]*Department of Mathematical Sciences, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA*

[2]*Department of Physics, , New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA*

[3]*Otto York Department of Chemical Engineering, New Jersey Institute of Technology, University Heights, Newark, NJ 07102, USA*

Spectroscopic analysis in the terahertz frequency range, providing characteristic "signatures" for explosive and non-explosive materials, is proposed as an efficient and powerful tool for explosive identification. It is demonstrated that spectral responses of materials can be used as fingerprints that distinguish RDX from other materials, even when measurements at only a limited number of frequencies are available. RDX detection comprises two stages: (i) a spline expansion that smoothes measured, noisy spectra, which is then differentiated for spectrum shape estimation, and (ii) a probability calculation, using Bayesian logistic regression, of whether a material belongs to the RDX class or not. The performance of the proposed detector is evaluated through application to spectra of RDX and several common materials. The detector is characterized by the desirable properties of a high probability of correct detection and a low probability of false alarm.

Section I: Introduction

The newly urgent need for extensive, reliable, and fast screening for the identification of explosives stemming from public safety and homeland security concerns has stimulated the search for efficient, quantitative detectors and classifiers. Terahertz (THz) spectroscopy has been proposed as a valuable tool to this end [1, 2, 3, 4, 5, 6, 7, 8]. It is a particularly useful approach because lethal agents have distinct spectral signatures in the THz frequency range that can serve as "fingerprints" for their identification.

In [8], the presence or absence of agents such as RDX (cyclotrimethylenetrinitramine) was detected by a least-squares distance calculation between measured and reference reflectance spectra in the THz range. The approach was supported by analysis of RDX spectra measured in the lab, clearly indicating a common spectroscopic behavior across several different scans of RDX. Classification of materials and identification of RDX was straightforward when reflectance level was taken into account.

True reflectance level, however, being susceptible to attenuating barriers and calibration factors, cannot be accurately linked to a specific material and cannot form an identifying signature in practice. Removing the true reflectance level results in identification based only on shape of spectroscopic curves across frequency. Preliminary work in [8] demonstrated that relative levels can indeed be employed for RDX identification, with an increase in probability of false alarm being the penalty of identification based only on shape.

Explosive identification was pursued in [6, 7], employing reflectances in the THz range as well. Transforming reflectance to absorption, classification is possible by matching the absorption peaks to known material characteristics. The transform, involving differentiation, has

the advantage of eliminating constants, and so the arbitrary level difficulties described above do not arise.

In this work, we extend the approaches of [6, 7, 8] and explore how a priori known shapes of spectral signatures of materials in the THz range can be optimally utilized for the implementation of a robust explosive detector-classifier when measurements at only a few frequencies are available. In our approach, penalized splines and differentiation are used to extract shape information from the data, similar to the approach employed in [9] for clustering gene expression data. Using Bayesian logistic regression, which was successfully used in phoneme classification applied to signatures similar to ours [10], we construct a classifier that is applied to the spline-smoothed spectra shapes and successfully distinguishes RDX from other materials in several real data experiments. It was of particular interest to us to study the performance of the proposed classifier-detector as a function of the number of frequencies at which reflectance measurements are available. This performance is quantified through extensive testing by employing measurements at subsets of frequencies.

The paper is organized as follows. Section II describes the data acquisition process. Section III discusses how splines are used for shape extraction from THz reflection spectra. Section IV presents Bayesian logistic regression as it applies to explosive detection. Section V demonstrates the performance of the developed detector-classifier on RDX detection employing sets of spectra measured in the lab and a Receiver Operating Characteristic (ROC) performance evaluation process. Conclusions are presented in Section VI.

Section II: Data acquisition

THz transmission and reflection spectra are measured using a Picometrix T-Ray 2000 THz time-domain spectrometer (THz-TDS). With the THz-TDS, the time-domain THz pulse is measured.

The roughly 80ps waveform is converted to the frequency domain via a Fourier transform. The usable frequency band is typically 0.1-2THz. Details of the THz-TDS method can be found in [11, 12].

The probing THz radiation is focused using silicon lenses to roughly a 3mm spot size on the sample. Reflection measurements from C4 explosives (RDX mixed with a plastique matrix) are performed on optically thick samples ensuring that the reflected THz waveform is due only to specular reflection from the front surface of the sample. Transmission measurements on thin samples reveal a pronounced absorption peak near 0.8THz [13]. In reflection, the presence of the absorption peak appears as a change of reflectance near 0.8THz. The reflectance is measured at multiple locations in order to sample the variation in measured reflectance due to material inhomogeneity, changes in surface topology.

Section III: A spline approach to modeling reflection spectra in the THz range

As noted in [10, 14], data preprocessing is a powerful tool for improving the performance of a learning algorithm. To facilitate the explosive detection procedure, we preprocess data by extracting the shape information of the reflectance spectra in the THz frequency range. For each material of interest, reflectance measurements are made in a specified frequency range; data consist of pairs $\{(w_k, s_k)\}$, $k = 1, 2, ..., K$, where w_k is a frequency and s_k is a measurement of reflectance at that frequency. The pair can be represented as:

$$s_k = f(w_k) + \varepsilon_k, \ \mathrm{E}(\varepsilon_k) = 0, \tag{1}$$

where ε_k is a random error and $\mathrm{E}(\varepsilon_k)$ is its expectation; experimentation has shown that a zero-mean Gaussian model for ε_k is a reasonable assumption. The goal is to estimate function f from data pairs $\{(w_k, s_k)\}$, $k = 1, 2, ..., K$, in a way that facilitates the extraction of shape information from the THz spectrum. Finding a functional expression for THz reflectance spectra is here pursued via penalized splines.

A spline $f(w)$ may be expressed as follows:

$$f(w) \approx \sum_{j=1}^{M} \beta_j \varphi_j(w), \tag{2}$$

where $\{\varphi_j\}$, $j = 1, 2, ..., M$, form a set of basis functions, and β_j, $j = 1, 2, ..., M$, are coefficients that combine the basis functions to build *f*. In this paper, we use a truncated power basis of degree 2:

$$1,\ w,\ w^2,\ (w-\xi_1)_+^2,\ (w-\xi_2)_+^2,\ ...,\ (w-\xi_{M-3})_+^2,$$

where $\xi_1,\ \xi_2, ...,\ \xi_{M-3}$ are spline knots, and $()_+ = \dfrac{|()| + ()}{2}$. Selecting second degree splines permits a continuous first derivative of the functional expression for the spectrum, necessary for shape estimation. Furthermore, second-degree splines enable us to fit a model using only a limited number of THz frequencies at which measurements are available.

The spline of Equation 2 may be expressed as

$$f(w) \approx \beta_1 + \beta_2 w + \beta_3 w^2 + \beta_4 (w-\xi_1)_+^2 + ... + \beta_M\ (w-\xi_{M-3})_+^2. \tag{3}$$

We focus on penalized splines to avoid overfitting by following random fluctuations in the data. According to [15], a penalized spline is defined as

$$\sum_{j=1}^{M}\beta_j\varphi_j(w) = \beta^T\Theta(w), \tag{4}$$

where $\Theta(w) = [\varphi_1(w), \varphi_2(w), ..., \varphi_M(w)]^T$, and

$$\beta = \underset{\beta}{\arg\min}\{\sum_{i=1}^{K}(s_i - \beta^T\Theta(w_i))^2 + \lambda^{2p}\beta^T D\beta\}, \tag{5}$$

where $p = 2$ (the degree of the splines), and $\lambda > 0$ is a Lagrange multiplier serving as a regularization parameter. D is an $M \times M$ diagonal matrix with the following elements: $D = diag[0,0,0,1,...,1]$. The first three diagonal elements of D are zeros, filtering through the first three components of β without restrictions. The remaining ones in matrix D penalize the coefficients of $(w-\xi_1)_+^2, (w-\xi_2)_+^2, ..., (w-\xi_{M-3})_+^2$ in Equation 2. Penalization, or regularization, produces estimates of β that provide reasonably small least-squared errors but do not model sharp data features that are results of noisy perturbations.

Parameter λ controls the degree of regularization. Large values of λ make the fit smoother; lower values allow more flexibility and, hence, a closer fit to the data. Smoothing imposed by regularization makes it possible to extract shape information from the data while avoiding fluctuations of THz spectra caused by noise; in essence, smoothing denoises the THz measurements. Parameter λ is readily selected for each experiment using generalized cross-validation [15, 16].

Let C be a $K \times M$ design matrix,

$$C = \begin{bmatrix} \varphi_1(w_1) & \varphi_2(w_1) & ... & \varphi_M(w_1) \\ \varphi_1(w_2) & \varphi_2(w_2) & ... & \varphi_M(w_2) \\ . & . & . & . \\ \varphi_1(w_K) & \varphi_2(w_K) & ... & \varphi_M(w_K) \end{bmatrix}. \quad (6)$$

Using *C,* we evaluate coefficients β necessary to construct *f(w)* of Equation 2 as follows [10, 15]:

$$\beta = (C^T C + \lambda^{2p} D)^{-1} C^T s\,,\ s = (s_1, s_2, ..., s_K)^T. \quad (7)$$

For all numerical experiments described below we employed MATLAB spline routines developed in [15].

Figure 1 illustrates three RDX reflectance spectra sampled at thirteen equally spaced frequencies ($K = 13$) between 0.6 and 0.9 THz and their spline approximations; the spectra are measured in the lab with the T-Ray system. These are superimposed on a mean RDX reflectance spectrum also employed in [8], which is in good agreement with the reference RDX reflectance spectrum of [6]. As previously discussed and also demonstrated by the figure, the spline representation has a smoothing effect on the measured spectra, which is critical for successful classification [10]; spline representations closely follow the shape of the mean RDX spectrum. In a similar fashion to Figure 1, Figure 2 depicts THz spectra of cloth and corresponding approximations with splines. These three spectra (as those from several other non-RDX materials) appear relatively uniform with frequency.

The spline analysis described in this section is applied to THz spectroscropic data in a stage preceding classification. The preprocessing stage steps are summarized as follows:

Preprocessing algorithm:

1. Fit the data with splines after calculating spline coefficients β_j.

2. Estimate spline derivatives to remove absolute level component and retain shape information:

$$\gamma^{(i)} = f^{'}(w_i) = \sum_{j=1}^{M} \beta_j \varphi_j'(w_i) = \\ = \beta_2 + 2\beta_3 w + 2\beta_4 (w - \xi_1)_+ + ... + 2\beta_M (w - \xi_{M-3})_+ . \quad (8)$$

Derivatives will be used as inputs to the classifier.

Section IV: Spectra classification using Bayesian logistic regression

The classification process consists of training and testing stages. The training data set consists of n data sets (t_r, γ_r), $r = 1,...,n$, where $\gamma_r = (\gamma_r^{(1)}, \gamma_r^{(2)}, ..., \gamma_r^{(K)})$ is a vector generated by the *preprocessing algorithm* described in Section III; for example, if we consider a 13-point THz spectrum (*K = 13*) at frequencies $(w_1, w_2, ..., w_{13})$, γ_r is a 13-dimensional vector of slopes of tangent lines to function $f(w)$ of Equation 2 at these 13 frequencies. Quantities t_r are target values that show whether the training data vector belongs to a specific category or not. Targets t are defined as:

$$t = \begin{cases} 1, \text{ if material is RDX;} \\ 0, \text{ otherwise.} \end{cases}$$

Binary logistic regression [17, 18] is employed here for classification; specifically, software *BBR*, developed by Genkin et al. [18, 19], was used. The approach calculates probability $P(C \mid \gamma)$ that an observation γ, $\gamma = (\gamma^{(1)}, \gamma^{(2)}, ..., \gamma^{(K)})$, belongs to class C (in our case, *C*="material is RDX"). This probability is modeled as a *logit* of the linear combination $\alpha_0 + \sum \alpha_k \gamma^{(k)}$, therefore:

$$P(C \mid \gamma) = \frac{\exp(\alpha_0 + \sum \alpha_k \gamma^{(k)})}{1 + \exp(\alpha_0 + \sum \alpha_k \gamma^{(k)})} \tag{9}$$

More specifically, we employ *Bayesian* logistic regression, estimating modes of regression coefficients α via a coordinate descent algorithm [18, 20]. For n data patterns in the training set **T**, the posterior probability distribution $p(\alpha|\mathbf{T})$ of α given data **T** may be expressed as follows:

$$p(\alpha|\mathbf{T}) \propto \left(\prod_{r=1}^{n} \left(\left(\frac{\exp(\alpha_0 + \sum \alpha_k \gamma_r{}^{(k)})}{1 + \exp(\alpha_0 + \sum \alpha_k \gamma_r{}^{(k)})} \right)^{t_r} \left(1 - \frac{\exp(\alpha_0 + \sum \alpha_k \gamma_r{}^{(k)})}{1 + \exp(\alpha_0 + \sum \alpha_k \gamma_r{}^{(k)})} \right)^{1-t_r} \right) \right) p(\alpha), \tag{10}$$

where $p(\alpha)$ is a prior probability distribution on regression coefficients. All prior distributions on individual regression coefficients α_k are assumed to be independent; the overall prior $p(\alpha)$ is thus a product of individual priors $p(\alpha_k)$. Gaussian prior distributions are used throughout this work. Modes of coefficients α_k are obtained by maximizing the distribution of Equation 10 over α.

Modal estimates of coefficients α determined during the training phase are then used for the estimation of probability of class membership (Equation 9) for each vector from the testing set. It is decided that a tested material belongs to the "RDX" class when the calculated probability exceeds a certain threshold, which can be determined by specifications on probabilities for correct detection and false alarm.

The maximization of the posterior distribution of Equation 10 rather than the likelihood for the estimation of coefficients α is the process that stabilizes the classifier's performance, enabling it to be successful in generalization. The variance of prior distribution $p(\alpha)$ is here the

regularization parameter, ensuring that the processor learns the training data well, but not at a level that would limit the classifier to recognizing and correctly identifying only the training vectors. Variance selection for the employed priors is performed with an L-curve analysis [21].

Typically, in classification approaches a principal component analysis of the data precedes the actual classification task [9]; the goal of this analysis is to identify uncorrelated data components that contribute information that is useful to the decision process. Instead of requiring this additional stage, Bayesian logistic regression includes such an analysis within the actual classification process by calculating modes of coefficients α of the data vectors (derivatives of splines at a set of frequencies) via maximization of Equation 10. When data components at specific frequencies are less informative, the modes of the corresponding coefficients are shrunk towards zero and play a smaller role in assigning the data vector to a class.

Section V: Spectra classification experiments

Classification was conducted with 713 THz spectra consisting of 13 data vectors that were utilized in [8], 500 additional data vectors of RDX, aluminum foil, cardboard, cloth, leather, plastic, plexiglass, wood, flour, and sugar (50 measurements of each material were taken), 100 measurements from a different, thicker RDX sample, and 100 measurements from the second RDX sample placed behind a nylon barrier. The original 13 spectra consisted of six measured spectra of RDX reflectances and one of pine wood, plexiglass, paper, cardboard, aluminum foil, skin, and plastic. To approximate realistic detection conditions, we concentrated on material identification using a limited number of THz frequencies.

We conducted the following classification experiments:

Experiment 1: Leave-one-out classification for each of the 13 reflectances from [8]. In the leave-one-out experiment, twelve spectra are used for training the Bayesian logistic regression classifier; the classifier is then tested on one remaining spectrum, not encountered in the training stage, to determine whether it belongs to the RDX class. The experiment is repeated 13 times, each time holding off one pattern for testing. Tables 1, 2, and 3 summarize the results of the experiment considering thirteen, seven, and five measurements at equally spaced frequencies between 0.6 and 0.9 THz.

Experiment 2: Detection of RDX/non-RDX materials using 713 reflectance measurement vectors from aluminum foil, cardboard, cloth, leather, plastic, plexiglass, wood, flour, sugar (50 spectra per material), the thirteen vectors of Experiment 1, 50 reflectance spectra from an RDX pellet, 100 spectra from a second, thicker RDX pellet, and 100 spectra from the latter RDX sample placed behind a barrier. Combining data from the two different RDX samples and including measurements from the second one behind a nylon barrier generates a diverse data set. Reflectance spectra from the new RDX pellet and their spline approximations are shown in Figures 3 and 4. These spectra are similar in shape to those from the first RDX sample displayed in Figure 1. All spectra demonstrate relatively stable reflectance values between 0.6 and approximately 0.77 THz; reflectance decreases after 0.77 THz. The measurements of Figure 4, although similar in general behavior to the spectra from the same RDX sample illustrated in Figure 3, exhibit lower levels, because of the presence of the barrier.

A set of 350 vectors (randomly selected from the full data set) were used for training; the remaining 363 smoothed and differentiated spectra were employed in testing. ROC curves were generated demonstrating classification performance with 13 and five-frequency data (Figures 6 and 7); results for seven frequencies are very similar and are not displayed.

Experiment 3: Experiment 2 was repeated, but, this time, all vectors from the training set that corresponded to reflectance data from the second RDX sample (with and without a barrier) were omitted from the training set, simulating a realistic situation where measurements of reflectance from unknown materials behind barriers are encountered. The classifier was then tested on the same data set as the one used in Experiment 2. Figures 8 and 9 present ROC curves for this experiment.

The first experiment demonstrated that the spectral shape of RDX reflectance is, indeed, a valuable feature for RDX detection and material classification in explosive or benign agents. A simple least-squares approach applied to the same data in [8] misclassified paper and aluminum foil, when normalized signatures were used. The new proposed algorithm correctly maps 12 materials in the dataset from [8] either to "RDX" or 'nonRDX'. Aluminum foil is the only material that is still misclassified. The aluminum foil reflectance level is much higher than that of RDX (and all other tested materials), making level a salient feature of the foil spectroscopic signature. When level is removed (using either normalization in [8] or differentiation here), the retained shape of the aluminum foil spectrum is similar to that of RDX, because of the orientation of the tested foil sample, explaining the persistent classification error. Figure 5 displays normalized spectra of RDX, paper, and aluminum foil, illustrating the challenge in distinguishing between materials based on reflectance shape and demonstrating the similarity between normalized foil and RDX spectra.

The results discussed above were obtained by using data vectors with thirteen reflectance measurements between 0.6 and 0.9 THz. When reflectance measurements at fewer frequencies within the same interval were made available to the classifier, classification performance deteriorated with four misclassifications for seven and five frequencies. This behavior is

expected, because this experiment relies on a training set that is very limited. Imposing additional limitations on the classifier by removing frequency samples deprives the processor of crucial information and leads to performance degradation.

Results from Experiment 2 were also quite promising. The ROC curves were constructed by selecting different threshold values for the processor output and computing for each value the probability of correct RDX identification (probability of detection) and probability of incorrectly identifying benign materials as RDX (probability of false alarm). For low thresholds both probabilities are expected to be high (reaching the top right corner of the figure); for large threshold values, these probabilities are very small (close to the bottom left corner). It is of particular interest to observe the probability behavior for intermediate threshold values. The ideal situation is to have, for certain thresholds, a high probability of detection coupled with a low probability of false alarm (with the ROC curve approaching the top left corner of the plot).

Our ROC curves (Figures 6 and 7 for 13 and five-frequency data, respectively) indeed show a behavior close to the one just outlined. For the 13-frequency case (Figure 6) and several thresholds, probability of correct RDX identification reaches 96% with a low probability of false alarm (10%). Results are similar but slightly degraded for data vectors with five frequencies; this is expected since limited shape information is conveyed when employing only five spectral points.

On the ROC curves of Figures 6 and 7, a second set of curves has been superimposed. These curves have been generated with a classifier that decides on class membership employing a distance criterion. This processor is similar to the one implemented in [7, 8] and evaluates the distance (or “closeness”) between measured reflectance spectra and a mean RDX spectrum, obtained from the training set and used as a reference. The unknown level has been removed

from the data through a simple subtraction of level at the lowest frequency. The RDX vectors from the training set of Experiment 2 were here considered as a reference from where an RDX "signature" was extracted and compared to testing data from RDX and other materials. The testing set, identical to that used with the new processor, was classified according to distances between data and reference. This standard approach has a significantly inferior performance to that of the processor developed in this work, with ROC curves reaching lower probabilities of detection than the new processor for the same levels of false alarm.

A similar performance evaluation was carried out for Experiment 3, with the ROC curves displayed in Figures 8 and 9. A comparison between Figures 8 and 6 shows that the new processor has been only minimally affected by the reduction in the number and diversity of RDX data vectors in the training stage, demonstrating that the processor has the ability to generalize, as also indicated from Experiment 1. This time, however, the minimum distance classifier is almost identical in terms of performance to the new processor. This improvement in the minimum distance results from Experiment 2 to Experiment 3 is attributed to the construction of the reference ("signature") RDX data vector to which all data are compared. In Experiment 2, the reference is constructed from reflectance data from two RDX pellets (including measurements from the second pellet behind a nylon barrier). This signature, providing less clear shape information than the reference vector constructed from RDX reflectance data from a single pellet, raises probability of false alarm levels, deteriorating detection performance of the minimum distance classifier in Experiment 3.

Figure 9, showing detection results for the five-frequency case, demonstrates that the new processor performs well in RDX detection and is superior to the distance classifier; for fixed, low probability of false alarm values, the new classifier produces a probability of detection that is

significantly higher to that of the minimum distance classifier (but the gap in performance is still reduced from that in Experiment 2).

Section VI: Conclusions

A method is developed for processing reflectance measurements from materials in the THz frequency range for explosive agent detection. Our technique is applied to measured reflectance data from various common materials including RDX, first providing a functional approximation to each spectrum and, subsequently, evaluating the derivative of the functional approximation with respect to frequency. The first step produces a smooth representation of a typically noisy set of measurements; the second step removes the absolute reflectance level but retains spectrum shape. After preprocessing, derivatives of spectra representations are presented to a Bayesian logistic regression classifier for training and testing. The classifier is trained to assign data vectors to an "RDX" or "non-RDX" class. Conducting three classification experiments, our work confirms previous findings [1, 2, 3, 4, 6, 7, 8, 12, 13] that spectroscopic information in the THz range is a powerful feature for material identification. The classifier is characterized by the desirable properties of high probability of correct RDX detection and low probability of false alarm. To simulate realistic test cases, we presented the classifier with several data sets (including reflectance measurements from an RDX sample behind a barrier), and compared the method to other conventional classification approaches. We observed consistently good classification performance of the proposed explosive identification technique, which is shown to be generally superior to a conventional minimum distance classifier. The processor performs well, even when measurements at only a few (five) frequencies are available.

Acknowledgment

The authors gratefully acknowledge the support of the US Army.

References

1. A. Bandyopadhyay, A. Sengupta, R.B. Barat, D.E. Gary, Z.-H. Michalopoulou, and J.F. Federici, “Artificial neural network analysis in interferometric Terahertz imaging for detection of lethal agents,” International Journal of Infrared and Millimeter Waves, vol. 27, no. 8, 1145-1158 (2006).
2. M. K. Choi, A.D. Bettermann, and W. Van derWiede, “Potential for detection of explosive and biological hazards with electronic terahertz systems,” Phil. Trans. R. Soc. Lond. A, vol. 362, 337-349 (2004).
3. J.F. Federici, D. Gary, R. Barat, and Z.-H. Michalopoulou, “Detection of Explosives by Terahertz Imaging,” in Counter-Terrorism Detection Techniques of Explosives, Jehuda Yinon, ed., pp. 323-366, (Elsevier, 2007).
4. J.F. Federici, B. Schulkin, F. Huang, D. Gary, R. Barat, F. Oliveira, and D. Zimdars, “THz imaging and sensing for security applications - Explosives, Weapons, and Drugs,” Semicond. Sci. Technol., vol. 20, S266-S280 (2005).
5. M.C. Kemp, P.F. Taday, B.E. Cole, J.A. Cluff, A.J. Fitzgerald, and W.R. Tribe, “Security applications of Terahertz technology,” in Terahertz for Military and Security Applications, R. J. Hwu and D. L.Woolard, eds., Proc. SPIE 5070, 44-52 (2003).
6. H.B. Liu, Y. Chen, G.J. Bastiaans, and X.-C. Zhang, “Detection and identification of explosive RDX by THz diffuse reflection spectroscopy,” Optics Express, vol. 14, issue 1, 415-423 (2006).

7. H. Zhong, A. Redo-Sanchez, and X.-C. Zhang, “Identification and classification of chemicals using terahertz reflective spectroscopic focal-plane imaging system,” Optics Express, vol. 14, issue 20, 9130-9141 (2006).
8. I. Zorych, A. Sinyukov, Z.-H. Michalopoulou, R. Barat, D. Gary, and J.F. Federici, “Explosive identification with Terahertz spectroscopy: a model based approach,” in *Proceedings of SAFE 2007*, pp. 114-117 (2007).
9. S. Dejean, P.G. Martin, A. Baccini, and P. Besse, “Clustering Time-Series Gene Expression Data Using Smoothing Spline Derivatives,” EURASIP Journal on Bioinformatics and Systems Biology, ID 70561, 10 pages (2007)
10. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, (Springer, 2001).
11. D. M. Mittleman, “Terahertz Imaging,” in *Sensing with Terahertz Radiation*, D. Mittleman, ed., pp. 117-153, (Springer, 2003).
12. D. M. Mittleman, M. Gupta, R. Neelamani, R.G. Baraniuk, J.V. Rudd, and M. Koch, “Recent advances in Terahertz imaging,” Appl. Phys. B., vol. 68, 1085-1094 (1999).
13. F. Huang, B. Schulkin, H. Altan, J. Federici, D. Gary, R. Barat**,** D. Zimdars, M. Chen, and D. Tanner, “Terahertz Study of 1,3,5-Trinitro-s-triazine (RDX) by Time Domain Spectroscopy and FTIR,” Appl. Phys. Lett., vol. 85, 5535-5537 (2004).
14. J. O. Ramsay and B.W. Silverman, *Functional Data Analysis,* (Springer, 2005).
15. D. Ruppert, M.P. Wand, and R.J. and Carroll, *Semiparametric Regression* (Cambridge University Press, 2003).
16. G. Wahba, “A Comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem,” Annals of Statistics, vol. 13, 1378-1402 (1985).

17. C. M. Bishop, *Pattern Recognition and Machine Learning*, (Springer, 2006).

18. A. Genkin, D. Lewis, and D. Madigan, "Large-scale Bayesian logistic regression for text categorization," *Technometrics*, vol. 49, num. 3, 291-304 (2007).

19. A. Genkin, D. Lewis, D. Madigan, "BBR: Bayesian Logistic Regression Software, User Guide and Software," http://www.stat.rutgers.edu/~madigan/BBR.

20. T. Zhang and F. Oles, "Text categorization based on regularized linear classifiers," Information Retrieval, vol. 4, issue 1, 5-31 (2001).

21. R. C. Aster, B. Borchers, and C. H. Thurber, *Parameter Estimation and Inverse Problems* (Elsevier Academic Press, 2005).

Table 1. Classification results: leave-one-out performance; spectral measurements were available at thirteen frequencies.

Number of frequencies =13

Frequencies (THz): 0.600, 0.625, 0.650, 0.675, 0.700, 0.725, 0.750, 0.775, 0.800, 0.825, 0.850, 0.875, 0.900

Material	Predicted Class	Predicted probability to be RDX	Prediction correct (Yes/No)
Plexiglass	nonRDX	0.048	Yes
Wood	nonRDX	0.262	Yes
Cardboard	nonRDX	0.022	Yes
Plastic	nonRDX	0.018	Yes
Paper	nonRDX	0.382	Yes
Aluminum foil	RDX	1.000	No
Skin	nonRDX	0.027	Yes
RDX1	RDX	0.913	Yes
RDX2	RDX	0.854	Yes
RDX3	RDX	0.654	Yes
RDX4	RDX	0.948	Yes
RDX5	RDX	0.928	Yes
RDX6	RDX	0.952	Yes

Table 2. Classification results: leave-one-out performance; spectral measurements were available at seven frequencies.

Number of frequencies =7 Frequencies(THz): 0.600, 0.650, 0.700, 0.750, 0.800, 0.850, 0.900			
Material	Predicted Class	Predicted probability to be RDX	Prediction correct (Yes/No)
Plexiglass	nonRDX	0.087	Yes
Wood	RDX	0.900	No
Cardboard	nonRDX	0.051	Yes
Plastic	nonRDX	0.050	Yes
Paper	RDX	0.762	No
Aluminum foil	RDX	1.000	No
Skin	nonRDX	0.060	Yes
RDX1	RDX	0.591	Yes
RDX2	RDX	0.850	Yes
RDX3	RDX	0.759	Yes
RDX4	nonRDX	0.352	No
RDX5	RDX	0.816	Yes
RDX6	RDX	0.927	Yes

Table 3. Classification results: leave-one-out performance; spectral measurements were available at five frequencies.

Number of frequencies =5 Frequencies(THz): 0.600, 0.675, 0.750, 0.825, 0.900			
Material	Predicted Class	Predicted probability to be RDX	Prediction correct (Yes/No)
Plexiglass	nonRDX	0.307	Yes
Wood	RDX	0.920	No
Cardboard	nonRDX	0.142	Yes
Plastic	nonRDX	0.132	Yes
Paper	nonRDX	0.083	Yes
Aluminum foil	RDX	1.000	No
Skin	nonRDX	0.157	Yes
RDX1	RDX	0.870	Yes
RDX2	RDX	0.780	Yes
RDX3	nonRDX	0.489	No
RDX4	nonRDX	0.487	No
RDX5	RDX	0.549	Yes
RDX6	RDX	0.806	Yes

Figure 1: Three measured RDX spectra at thirteen frequencies between 0.6 and 0.9 THz (blue lines with circles), corresponding approximations with splines (green solid lines), and a reference RDX spectrum (red dot-dashed line).

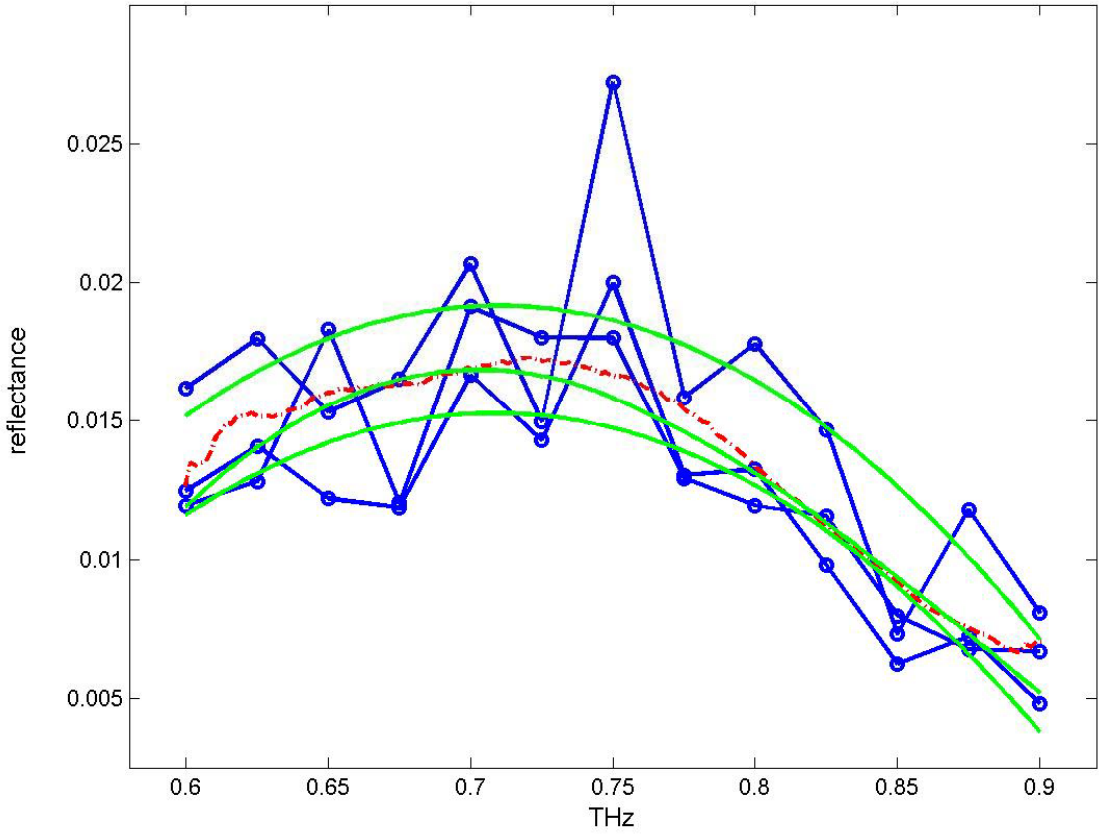

Figure 2: Spectra from plexiglass, wood, and skin at thirteen frequencies between 0.6 and 0.9 THz (solid lines with circles, asterisks, and diamonds, respectively) and corresponding approximations with splines (green solid lines).

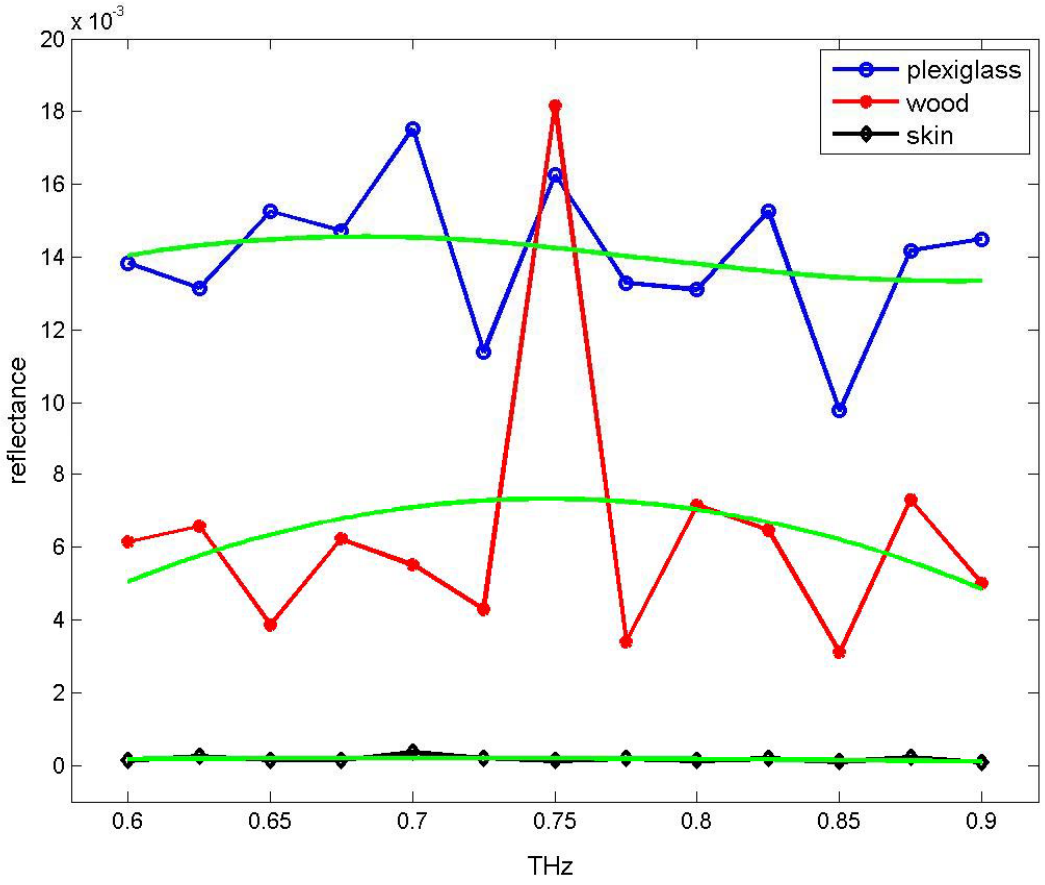

Figure 3: Three reflectance spectra from a thick RDX sample (solid lines with circles) and their spline approximations (solid lines).

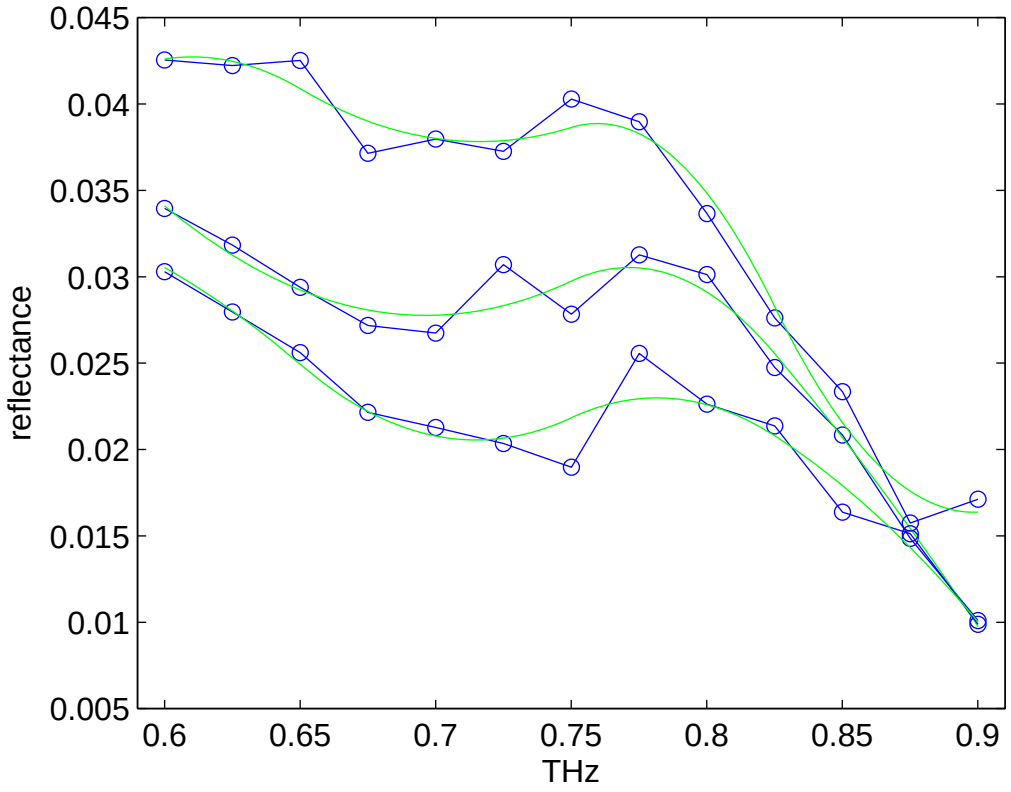

Figure 4: Three reflectance spectra from a thick RDX sample behind a nylon barrier (solid lines with circles) and their spline approximations (solid lines).

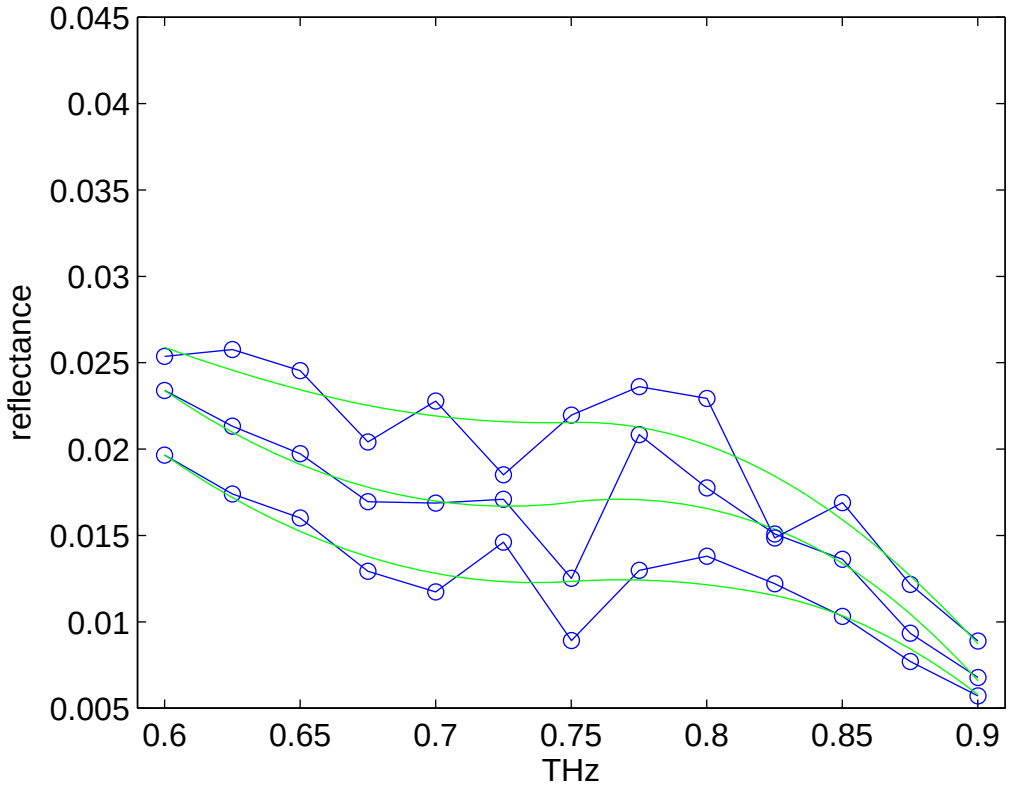

Figure 5: Normalized reflectance spectra: paper (dashed line), aluminum foil (dot-dashed line), and RDX (solid line).

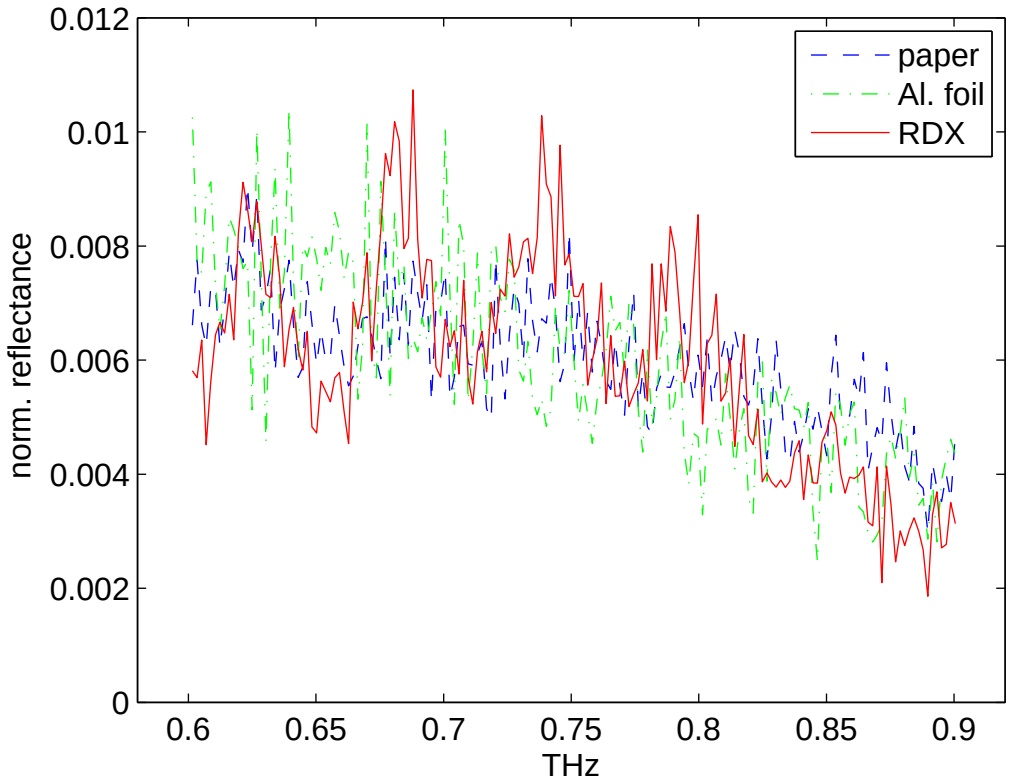

Figure 6. ROC curves for the Bayesian Binary Regression classifier (‘o’) and a minimum distance classifier (‘-.’) for Experiment 2 and 13 frequencies.

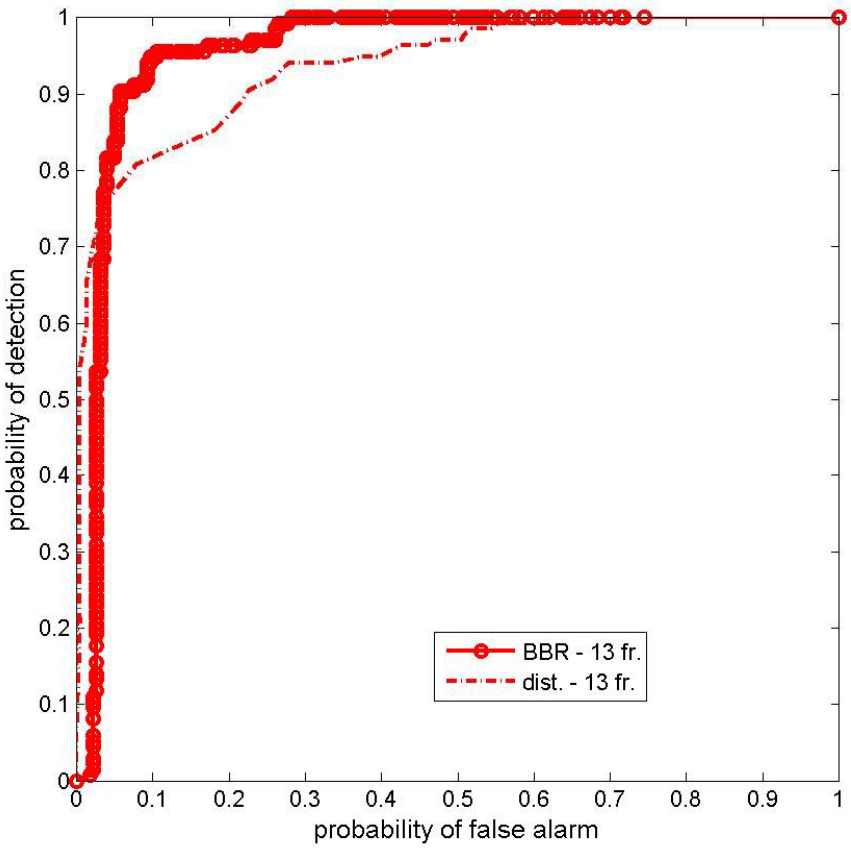

Figure 7. ROC curves for the Spline/Bayesian Binary Regression classifier (‘o’) and a minimum distance classifier (‘-.’) for Experiment 2 and five frequencies.

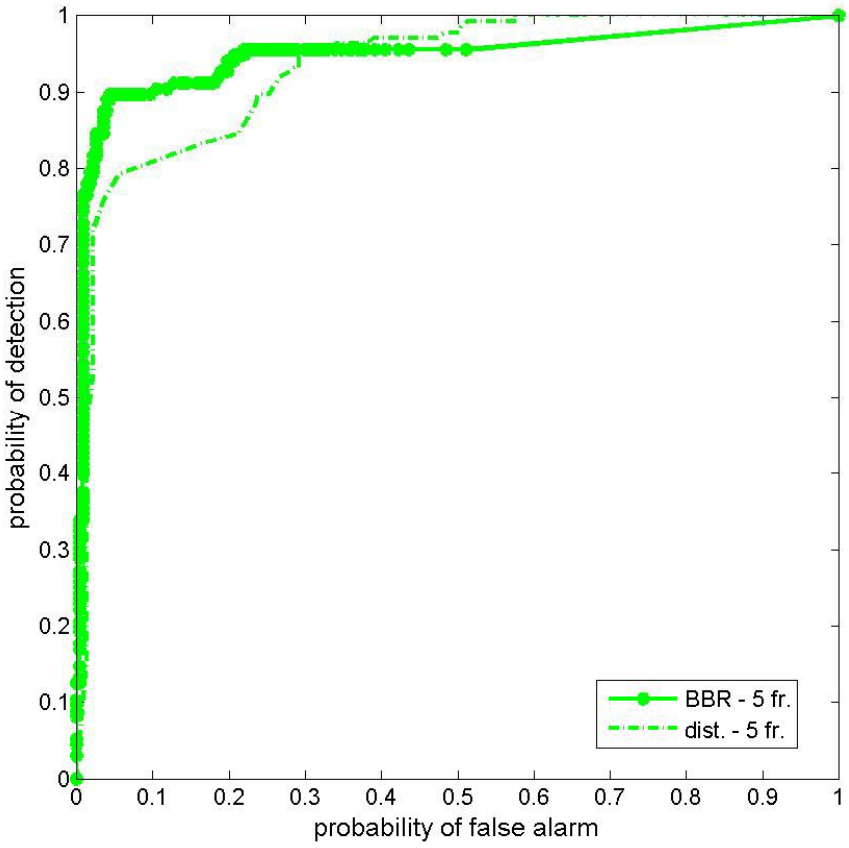

Figure 8. ROC curves for the Bayesian Binary Regression classifier (‘o’) and a minimum distance classifier (‘-.’) for Experiment 3 and 13 frequencies.

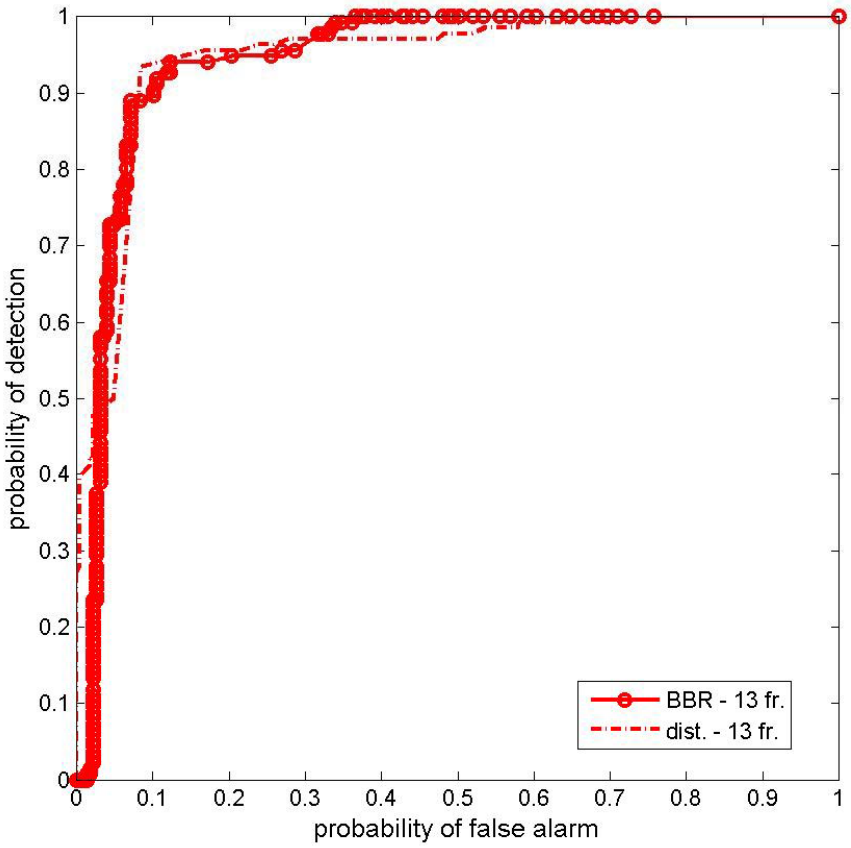

Figure 9: ROC curves for the Spline/Bayesian Binary Regression classifier (‘-o-’) and a minimum distance classifier (‘-.’) for Experiment 3 and five frequencies.