INFRASOUND SIGNAL SEPARATION USING INDEPENDENT COMPONENT ANALYSIS

Fredric M. Ham and Nizar A. Faour, Florida Institute of Technology
Joseph C. Wheeler, The Boeing Company

Sponsored by The Boeing Company
Contract No. 7M210007

ABSTRACT

An important element of monitoring compliance of the Comprehensive Nuclear-Test-Ban Treaty (CTBT) is an infrasound network. For reliable monitoring, it is important to distinguish between nuclear explosions and other sources of infrasound. This will require signal (event) classification after a detection is made. We have demonstrated the feasibility of using neural networks to classify various infrasonic events. However, classification of these events can be made more reliably with enhanced quality of the recorded infrasonic signals. One means of improving the quality of the infrasound signals is to remove background noise. This can be carried out by performing signal separation using Independent Component Analysis (ICA). ICA can be thought of as an extension of Principal Component Analysis (PCA). Using ICA, noise, and other events that are not of concern, can be removed from the signal of interest. This is not a *filtering* process, but rather a technique that actually *separates* out the background noise from the signal of interest, even if the signals have overlapping spectra. Therefore, not only is the signal of interest *recovered*, but so is the background noise. The higher fidelity signal of interest (compared to any one sensor channel signal from the infrasound array before separation) can be presented to an event classifier (e.g., a neural network), and the background noise can also be further scrutinized.

We show two examples of infrasound signal separation using ICA. The ICA is performed using a neural network approach, i.e., an unsupervised nonlinear PCA subspace learning rule. The first example involves artificially mixing three different infrasonic signals from three separate events using a random mixing matrix, these mixed signals are then used to recover the original event signals. The second example is in the true spirit of ICA, i.e., the separation is performed blindly. From four channels of an infrasound array, these four inputs are used in the ICA to separate two signals, i.e., one "signal" the other "noise." The mixing matrix is not known, however, the separated *signal* of interest is shown to be the infrasound signal of a volcano eruption, and the separated *noise* is shown to contain characteristics of a microbarom signal. Moreover, in spite of overlapping spectra between the output signals of the ICA, separation of the signals is possible.

Key Words: Infrasound, independent component analysis, signal separation, neural network, nonlinear principal component analysis, overlapping spectra.

OBJECTIVE

The proposed International Monitoring System (IMS), for verifying compliance of the Comprehensive Nuclear-Test-Ban Treaty (CTBT), includes an infrasound network that will be capable of detecting nuclear explosions. In the past, the primary interest in monitoring infrasound waves from nuclear explosions was to detect low-frequency signals associated with large (megaton) yield nuclear events. The main emphasis for monitoring compliance of the CTBT is the ability to detect and classify kiloton-size nuclear explosions. Infrasound signals from these events have a much higher frequency content than those from larger yield nuclear explosions. The infrasound signals received at a single station (multi-sensor array) can contain multiple events that can overlap in frequency and make reliable detection and classification of a nuclear explosion difficult. Therefore, the objective of this paper is to show how Independent Component Analysis (ICA) [1-4] can be used to separate infrasonic signals, in spite of overlapping spectra, recorded on a multi-sensor array. The ICA is performed using a neural network approach, specifically, an unsupervised nonlinear Principal Component Analysis (PCA) subspace learning rule. The purpose of the signal separation process is to separate the signal of interest from the background noise. It is assumed that there is one *dominant* event recorded along with background noise (which can consist of multiple, less dominant, components, including noise). The separated signal of interest will be higher quality compared to any one sensor channel signal from the infrasound array before separation. This high fidelity separated signal of interest can then be presented to an event classifier, e.g., a neural network [5, 6], and the background noise can also be further evaluated.

RESEARCH ACCOMPLISHED

A succinct presentation of ICA is given first followed by a brief discussion of how ICA can be carried out using neural networks. Finally, two examples of signal separation using a neural network ICA approach is presented.

Independent Component Analysis

Independent component analysis (ICA) can be thought of as an extension of PCA [7]. It is used primarily to separate unknown source signals from their linear mixtures, this is known as the blind source signal separation problem. The characteristics of the transmission channel does not have to be known to separate the source signals from a set of noisy observable (measured) signals. Moreover, even when the source signals have overlapping spectra, separation of the signals is possible. Blind source separation techniques can be applied to array processing, medical signal processing, communications, speech processing, and image processing, to name a few. The main difference between PCA and ICA is that instead of the uncorrelatedness property associated with standard PCA, in ICA the coefficients of the linear expansion of the data vectors must be mutually independent, or as independent as possible. What this means is that higher-order statistics [8, 9] must be used to determine the ICA expansion. In standard PCA, second-order statistics provides only decorrelation. Higher-order statistics are useful when dealing with non-Gaussian processes, non-minimum phase problems, colored noise, or even nonlinear processes [8]. Therefore, it is not surprising that nonlinearities must be used in the learning phase, even though the final input/output mapping is linear [10]. Neural network structures have been developed to perform ICA [10].

We assume that there exist q *zero-mean* wide-sense stationary source signals $s_1(k), s_2(k), \cdots, s_q(k)$ for $k = 1, 2, \cdots$ (the discrete time index or for images the pixels), that are scalar-valued and mutually statistically independent for each sample value k. The independence condition can be formally defined by stating that the joint probability density of the source signals is equal to the product of the marginal probability densities of the individual signals, i.e.,

$$p[s_1(k), s_2(k), \cdots, s_q(k)] = p[s_1(k)]p[s_2(k)]\cdots p[s_q(k)] = \prod_{i=1}^{q} p[s_i(k)] \tag{1}$$

The individual source signals are also assumed to be unknown (unobservable), however, we do have access to a set of h noisy linear mixtures of the unknown signals, $x_1(k), x_2(k), \cdots, x_h(k)$. These measured signals are given by

$$x_j(k) = \sum_{i=1}^{q} s_i(k) a_{ij} + n_j(k) \tag{2}$$

for $j = 1, 2, \cdots, h$, the elements a_{ij} are assumed to be not known, and $n_j(k)$ is additive measurement noise. We can now define the following, $\boldsymbol{x}(k) = \left[x_1(k) \quad x_2(k) \quad \cdots \quad x_h(k)\right]^T$, $\boldsymbol{x}(k) \in \Re^{h\text{x}1}$, $\boldsymbol{s}(k) = \left[s_1(k) \quad s_2(k) \quad \cdots \quad s_q(k)\right]^T$, $\boldsymbol{s}(k) \in \Re^{q\text{x}1}$ (the *source vector* consisting of the q independent components), and $\boldsymbol{A} = \left[\boldsymbol{a}_1 \quad \boldsymbol{a}_2 \quad \cdots \quad \boldsymbol{a}_q\right]$, $\boldsymbol{A} \in \Re^{h\text{x}q}$ (the *mixing matrix*), where the column vectors of $\boldsymbol{A}$ are the *basis vectors* of the ICA expansion. Equation (2) can now be written in vector-matrix form as

$$\boldsymbol{x}(k) = \boldsymbol{A}\boldsymbol{s}(k) + \boldsymbol{n}(k) = \sum_{i=1}^{q} s_i(k)\boldsymbol{a}_i + \boldsymbol{n}(k) \tag{3}$$

referred to as the ICA expansion. We will assume the mixing matrix $\boldsymbol{A}$ contains at least as many rows as columns ($h \geq q$), and it has full column rank, i.e., $\rho(\boldsymbol{A}) = q$ (i.e., the mixtures of the source signals are all different).

Independent Component Analysis Using Neural Networks

This discussion pertaining to the neural network approach for blind source separation using ICA follows the presentation by Karhunen *et al.* [10]. Figure 1 shows the basic neural architecture to perform the separation of source signals (i.e., estimate the independent components), and estimate the basis vectors of the ICA expansion, [i.e., estimate the column vectors of the mixing matrix $\boldsymbol{A}$ in (3)].

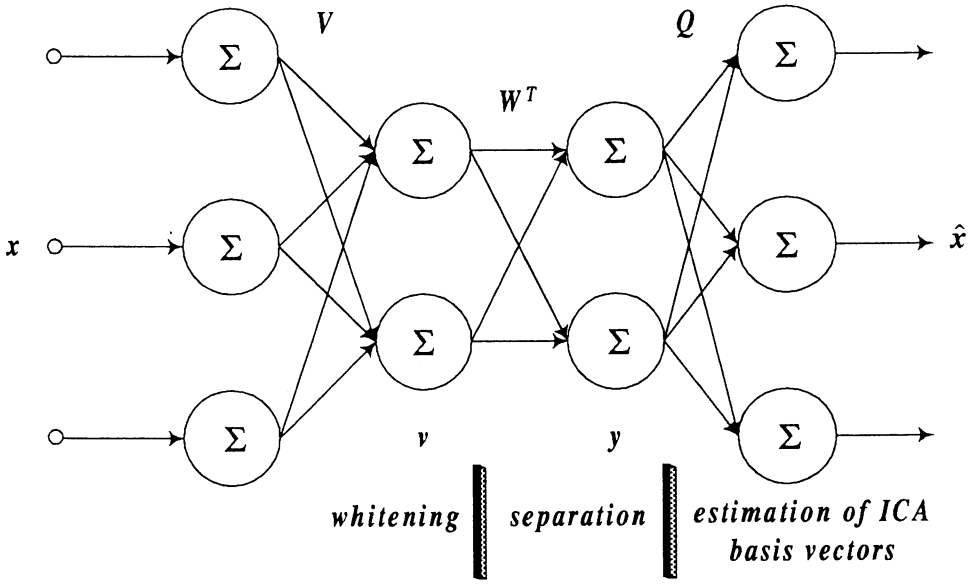

Figure 1. The ICA network. The three layers perform whitening, separation, and estimation of the basis vectors. The weight matrices that are necessary to determine are $\boldsymbol{V}$, $\boldsymbol{W}^T$, and $\boldsymbol{Q}$.

Prewhitening Process

The whitening process that proceeds the separation step (i.e., prewhitening) is a critical procedure. This process normalizes the variances of the observed signals to unity. In general, separation algorithms that use prewhitened inputs often have better stability properties and converge faster. However, whitening the data can make the separation problem more difficult if the mixing matrix $\boldsymbol{A}$ is ill-conditioned or if some of the source signals are relatively *weak* compared to the other signals [11, 12]. The input vectors $\boldsymbol{x}(k)$ are whitened by applying the transformation

$$\boldsymbol{v}(k) = \boldsymbol{V}\boldsymbol{x}(k) \tag{4}$$

where $\boldsymbol{v}(k)$ is the kth whitened vector and $\boldsymbol{V}$ is the whitening matrix. The whitening matrix can be determined in two ways: (i) using a batch approach, or (ii) neural learning. For the batch approach, PCA is used to determine the whitening matrix, it is given as

$$\boldsymbol{V} = \boldsymbol{D}^{-1/2}\boldsymbol{E}^T \tag{5}$$

where $\boldsymbol{V} \in \Re^{qxh}$, $\boldsymbol{D} = \text{diag}\left[\lambda_1 \quad \lambda_2 \quad \cdots \quad \lambda_q\right] \in \Re^{qxq}$, and $\boldsymbol{E} = \left[\boldsymbol{c}_1 \quad \boldsymbol{c}_2 \quad \cdots \quad \boldsymbol{c}_q\right] \in \Re^{hxq}$, with λ_i: ith largest eigenvalue of the covariance matrix $\boldsymbol{C}_x = \text{E}\left\{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right\} \in \Re^{hxh}$, and $\boldsymbol{c}_i$ for $i = 1, 2, \cdots, q$ are associated (principal) eigenvectors. Therefore, the transformation in (4) actually consists of two steps, i.e., compression and whitening. The compression step consists of selecting the proper value for q (the number of source signals). Therefore, the PCA described above for the whitening can also be used to select (i.e., estimate) the number of source signals (q) to be recovered (or the number of independent components) if the noise term $\boldsymbol{n}(k)$ in (3) is assumed to be zero-mean Gaussian white noise with covariance matrix $\text{E}\left\{\boldsymbol{n}(k)\boldsymbol{n}^T(k)\right\} = \sigma^2\boldsymbol{I}_h$. In the noise covariance matrix, σ^2 is the conjoint variance of the components of the noise vector $\boldsymbol{n}(k)$. The noise vector is assumed to be uncorrelated with the sources $s_i(k)$, for $i = 1, 2, \cdots, q$. Given these assumptions, the covariance matrix of the data vectors $\boldsymbol{x}(k)$ is given by

$$\text{E}\left\{\boldsymbol{x}(k)\boldsymbol{x}^T(k)\right\} = \sum_{i=1}^{q} \text{E}\left\{s_i^2(k)\right\}\boldsymbol{a}_i\boldsymbol{a}_i^T + \sigma^2\boldsymbol{I}_h \tag{6}$$

The q largest eigenvalues of the covariance matrix in (6), i.e., $\lambda_1, \lambda_2, \cdots, \lambda_q$, are some linear combination of the source signal powers $\text{E}\left\{s_i^2(k)\right\}$ added to the noise power σ^2. Therefore, the remaining $h - q$ eigenvalues correspond to only noise (theoretically these eigenvalues are equal to σ^2). The q largest *signal* eigenvalues will be distinctly larger than the remaining *noise* eigenvalues if the signal-to-noise ratio is large enough. In practice, the eigenvalues of the input covariance matrix are determined from the time-average of the covariance matrix over the available data vectors given by

$$\boldsymbol{C}_x \cong \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{x}_i(k)\boldsymbol{x}_i^T(k) \tag{7}$$

where N is the total number of input vectors. A stochastic approximation algorithm to *learn* the whitening matrix is given by

$$\boldsymbol{V}(k+1) = \boldsymbol{V}(k) - \mu(k)[\boldsymbol{v}(k)\boldsymbol{v}^T(k) - \boldsymbol{I}]\boldsymbol{V}(k) \tag{8}$$

where it is recommended to adjust the learning rate parameter according to

$$\mu(k) = \frac{1}{\dfrac{\gamma}{\mu(k-1)} + \left\|\boldsymbol{v}(k)\right\|_2^2}, \quad 0 < \gamma \le 1.0 \tag{9}$$

where $\left\|\bullet\right\|_2$ is the L_2 or Euclidean norm of a vector, and γ is the forgetting factor. When the whitening (orthogonal) transformation $\boldsymbol{V}$ is applied to the inputs as in (4), the resulting whitened outputs $\boldsymbol{v}(k)$ will possess the whiteness condition, i.e.,

$$\text{E}\left\{\boldsymbol{v}(k)\boldsymbol{v}^T(k)\right\} = \boldsymbol{I}_q \tag{10}$$

Separation Process

The separation process can be carried out by using many different methods [4, 11, 13]. Approximating contrast functions, maximized by separating matrices, have been developed [4]. Contrast functions typically require extensive batch computations using estimated higher-order statistics of the data and lead to very complicated

adaptive separation algorithms. However, it is sufficient to use the kurtosis (fourth-order cumulant) of the data. Another class of separation methods involves using neural networks to perform the separation of the source signals [12]. In Fig. 1, the second stage of the architecture is responsible for the separation of the whitened signals v. The linear separation transformation is given by

$$y(k) = W^T v(k) \tag{11}$$

where $W \in \Re^{qxq}$ ($W^T W = I_q$) is the separation matrix. Thus the separated signals are the outputs of the second stage, i.e., $\hat{s}(k) = y(k)$. An interesting observation is once the source signal $s(k)$ has been estimated, this means that the pseudo-inverse of A, i.e., A^+, must have been also "blindly" determined [refer to (3)].

One very straightforward neural learning method to determine the separation matrix is based on the nonlinear PCA subspace learning rule [14-16] given by

$$W(k+1) = W(k) + \mu(k)[v(k) - W(k)g\{y(k)\}]g\{y^T(k)\} \tag{12}$$

where $v(k)$ is the prewhitened input vector given in (4), and the function $g(\bullet)$ is a suitably chosen nonlinear function usually selected to be odd in order to ensure stability and for separation purposes. It is recommended that the learning rate parameter $\mu(k)$ be adjusted according to the adaptive scheme given in (9), with $v(k)$ replaced by $y(k)$. Also, for good convergence, it is best to select the initial weight matrix $W(0)$ to have as columns a set of orthonormal vectors. Typically, the nonlinear function $g(\bullet)$ is chosen as

$$g(t) = \beta \tanh(t / \beta) \tag{13}$$

where $g(t) = \dfrac{df(t)}{dt}$ and $f(t) = \beta^2 \ln[\cosh(t / \beta)]$, the logistic function. This is not an arbitrary choice for the nonlinearity in the learning rule of (12). It is motivated by the fact that when determining the ICA expansion *higher-order statistics* are needed. This can be seen by observing another neural learning rule to perform separation of unknown signals. This learning rule is called the *bigradient algorithm* [10, 17, 18] given by

$$W(k+1) = W(k) + \mu(k)v(k)g[y^T(k)] + \gamma(k)W(k)[I - W^T(k)W(k)] \tag{14}$$

where $\gamma(k)$ is another *gain* parameter, typically about 0.5 or one. This is a stochastic gradient algorithm that maximizes or minimizes the performance criterion

$$J(W) = \sum_{i=1}^{q} E\{f(y_i)\} \tag{15}$$

under the constraint that the weight matrix W must be orthonormal. The orthonormal constraint in (15) is realized in the learning rule in (14) in an additive manner. With the appropriate function $f(\bullet)$ in (15), the performance criterion would involve the sum of the fourth-order statistics (fourth-order cumulants) of the outputs, i.e., the *kurtosis* [8]. Therefore, the criterion would be either *minimized* for sources with a *negative kurtosis* and *maximized* for sources with a *positive kurtosis*. Source signals that have a *negative kurtosis* are often called *sub-Gaussian* signals and sources that have a *positive kurtosis* are referred to as *super-Gaussian* signals. In (15) the expectation operator would be dropped because we only consider instantaneous values. We now write the logistic function $f(t) = \ln[\cosh(t)]$ (for $\beta = 1$) in terms of a Taylor series expansion

$$f(t) = \ln[\cosh(t)] = t^2 / 2 - t^4 / 12 + t^6 / 45 - \cdots \tag{16}$$

The second-order term $t^2 / 2$ is on the average constant due to the whitening. The nonlinearity would then be given by $g(t) = \dfrac{df(t)}{dt} = \tanh(t) = t - t^3 / 3 + 2t^5 / 15 - \cdots$, and the cubic term will be dominating (an odd function) if the data are prewhitened.

Estimation of the ICA Basis Vectors

This is the last stage in Fig. 1. Two basic methods can be used to estimate the ICA basis vectors, or the column vectors of the mixing matrix A in (3). The first method is a "batch" approach where the estimate of A, i.e., $\hat{A}$, is given by

$$\hat{A} = ED^{1/2}W \tag{17}$$

where D is the eigenvalue matrix shown in (5), E has columns that are the associated eigenvectors shown in (5), and W is the separation matrix. The second method is a neural approach for estimating the ICA basis vectors. From Fig. 1, the last stage gives an estimate of the observed data as

$$\hat{x} = Qy \tag{18}$$

Comparing (18) with (3) for $n=0$ (i.e., $x = As$), we see that $Q = \hat{A}$ since $y = \hat{s}$. Therefore, the columns of the Q matrix are estimates of the columns of A, the ICA basis vectors. A neural learning algorithm can be derived from a representation error performance measure given by

$$J(Q) = \frac{1}{2}\|x - \hat{x}\|_2^2 = \frac{1}{2}\|x - Qy\|_2^2 \tag{19}$$

Taking a steepest descent approach given by $Q(k+1) = Q(k) - \mu\nabla_Q J(Q)$, the neural learning rule for estimating the ICA basis vectors is

$$Q(k+1) = Q(k) + \mu(k)[x(k) - Q(k)y(k)]y^T(k) \tag{20}$$

where $\mu > 0$ is the learning rate parameter that can be adapted during learning using (9) with $v(k)$ replaced by $Q(k)y(k)$.

Simulation 1

This example involves separating three different *artificially mixed* infrasonic signals. These infrasonic signals were recorded from a single station, 4-sensor (F-array), infrasound array in Windless Bight, Antarctica as separate events. The three signals are shown in Fig. 2(a), and are (i) an infrasonic signal from a volcano eruption at Galunggung, Java (recorded in 1982), (ii) a mountain associated wave originating from New Zealand (recorded in 1983), and (iii) and an internal atmospheric gravity infrasound wave (recorded in 1983). The three signals are *artificially* mixed using a random mixing matrix given by

$$A = \begin{bmatrix} 0.3050 & 0.9708 & 0.4983 \\ 0.8744 & 0.9901 & 0.2140 \\ 0.0150 & 0.7889 & 0.6435 \\ 0.7680 & 0.4387 & 0.3200 \end{bmatrix} \tag{21}$$

Therefore, four observed mixed signals are generated from $x(k) = As(k)$, for $k = 1, 2, \cdots, 768$, and are shown in Fig. 2(b). The eigenvalues of the estimated covariance matrix of the observed data, given by (7), are $\lambda_1 = 2.1346$, $\lambda_2 = 0.1976$, $\lambda_3 = 0.0434$, and $\lambda_4 = -3.7772\text{x}10^{-16}$. The fourth eigenvalue is considerably smaller than the first three. Therefore, only the first three largest ones need to be retained, and from (5) the whitening matrix $V \in \Re^{3x4}$ provides both whitening of the observed data and compression. So $h = 4$ and $q = 3$ (the number of source signals to be recovered). The nonlinear PCA neural learning rule in (12) is used to compute the separation matrix $W \in \Re^{3x3}$, and the nonlinearity $g(\bullet)$ is selected as the derivative of the logistic function as shown in (13) with $\beta = 1$. The initial weight (separation) matrix $W(0)$ is selected randomly, however, the columns of the matrix are constrained to be orthonormal. The learning rate adjustment scenario in (9) is used with $v(k)$ replaced by $y(k)$, and the forgetting factor is set at $\gamma = 0.9$. It required 250 training epochs for the neural network weights to converge. Fig. 2(c) shows the resulting three separated signals. Because we know what the source signals are, the correlation coefficient can be computed for each of the separated signals with respect to the known (actual) source signals. These correlation coefficients are shown in Fig. 2(c). The correlations of the

separated signals with respect to the actual source signals are almost perfect. The negative correlation coefficient indicates that a 180^{o} phase shift has occurred in the output of the ICA separation process.

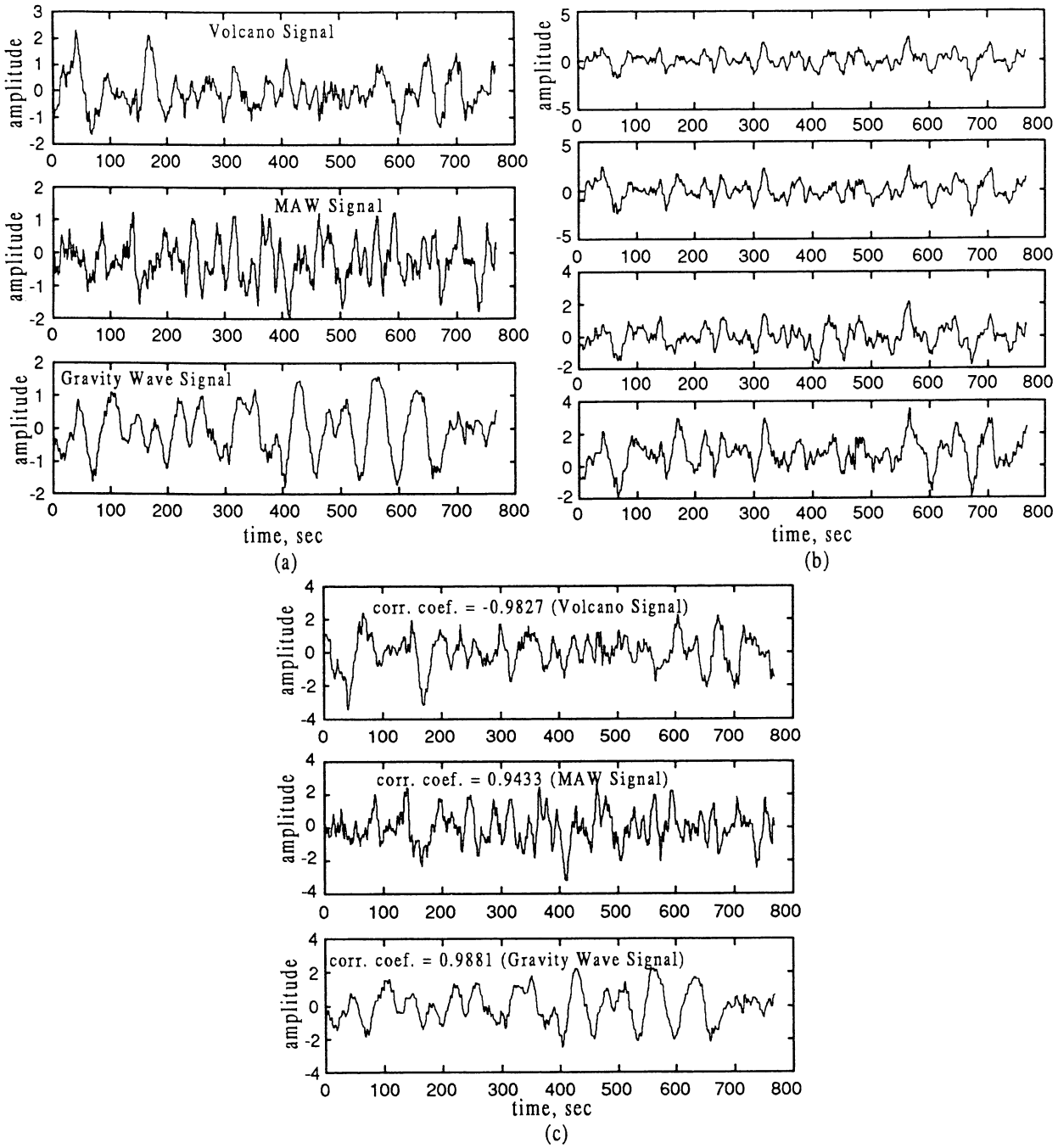

Figure 2. (a) Three original infrasonic source signals. (b) "Observed" mixed infrasound signals. (c) Separated infrasonic source signals using the nonlinear PCA subspace learning rule.

Simulation 2

The second example involves processing four infrasound signals from a large volcano eruption in Galunggung, Java in 1982. The signals analyzed were recorded from a single station, 4-sensor (F-array), infrasound array in Windless Bight, Antarctica. Fig. 3(a) shows the four recorded signals after beamforming [5] is applied to time align the signals to a common reference in the sensor array. The nominal sampling frequency is 1 Hz, and 590 time samples were retained from each signal record for analysis. It is assumed that the number of source signals is

two, i.e., a signal of interest and noise. Therefore, in the ICA, q=2 and n=4, the number of observed signals, i.e., the measured signals that are shown in Fig. 3(a) (which are the inputs to the ICA). The inputs are first prewhitened using the batch approach given in (4) and (5). Only the first two (largest) eigenvalues are retained along with the associated eigenvectors. Therefore, the whitening matrix $\boldsymbol{V} \in \Re^{2x4}$ computed using (5) provides both whitening of the observed data and compression. The nonlinear PCA neural learning rule in (12) is used to compute the separation matrix $\boldsymbol{W} \in \Re^{2x2}$, and the nonlinearity $g(\bullet)$ is selected as the derivative of the logistic function as shown in (13) with $\beta = 1$. The initial weight (separation) matrix $\boldsymbol{W}(0)$ is selected randomly, however, the columns of the matrix are constrained to be orthonormal. The learning rate adjustment scenario in (9) is used with the forgetting factor is set at $\gamma = 0.9$. It required 50 training epochs for the neural network weights to converge. Fig. 3(b) shows the two separated signals. The signal in the top graph appears to be the volcano infrasound signal and the signal in the bottom graph has the semblance of "noise." Figure 4(a) shows the first separated signals superimposed on the four input signals. The correlation coefficient computed for each input signal with the first separated signal is relatively high. Figure 4(b) shows the first separated signal superimposed on the average of the four input signals. In this case, the correlation coefficient is very high compared to the individual channel signals and the separated signal. This is due to reducing some of the noise in the signals from the simple averaging of the four input signals. Figure 5(a) shows the Power Spectral Density (PSD) of the four input signals, and Fig. 5(b) shows the PSD of the averaged input signal in the top graph and the PSD of the first separated signal in the bottom graph. The PSDs shown are an estimate of the respective signal's power spectrum based on the periodogram of the signals. Figures 4(a) & 4(b) give strong evidence that the first separated signal is the volcano signal. This is further evidenced in Fig. 5(b) by comparing the shape of the PSD of the average of the four input signals with the PSD of the first separated signal.

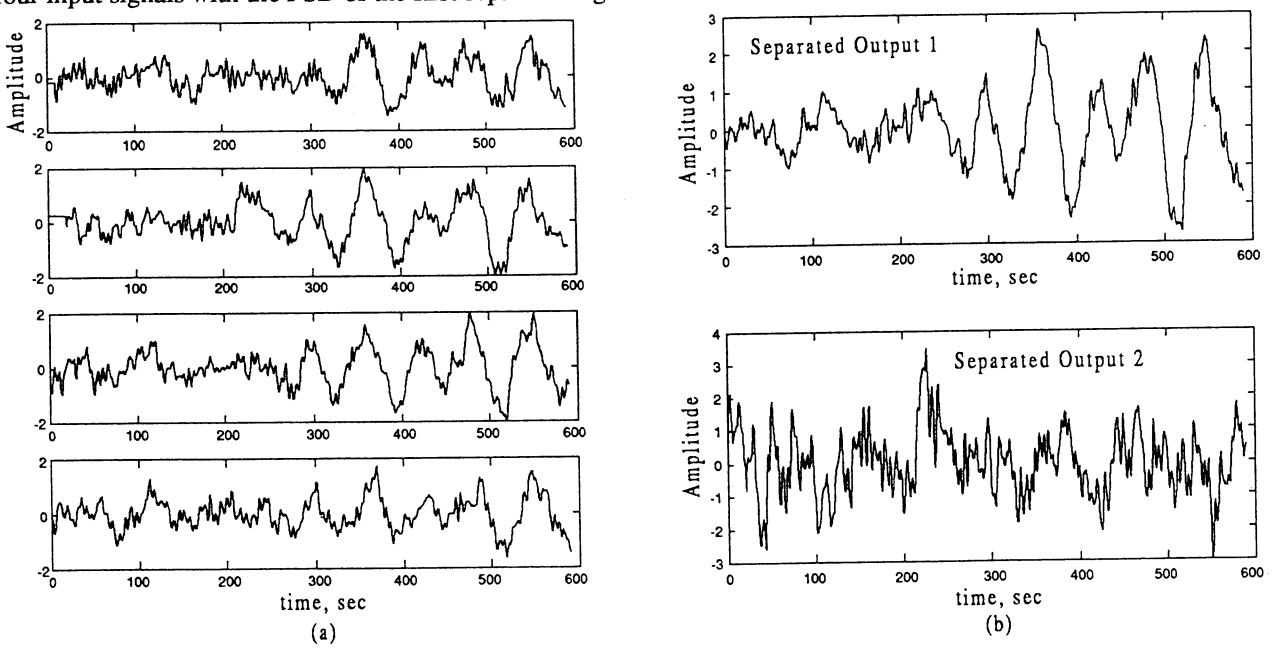

Figure 3. (a) Four volcano infrasound signals recorded on a four-sensor, single station, array in Windless Bight, Antarctica in 1982 (inputs to the ICA). (b) Two separated signals (outputs of the ICA).

Figure 6(a) shows the second ICA separated signal in the top graph and a typical microbarom signal in the bottom graph. The microbarom signal was recorded at Windless Bight, Antarctica on a T-array (3-sensor) infrasound array (the nominal sampling frequency is 4 Hz). In Fig. 6(b), the top graph is a rescaled version of the PSD of the first separated signal show in the bottom graph of Fig. 5(b), and bottom graph is the PSD of the microbarom signal. Figure 6(c) shows the PSD of the second separated signal in the top graph, and the PSD of the microbarom signal in the bottom graph, and Fig. 6(d) is a rescaled version of Fig. 6(c). Comparing the top and bottom graphs in Fig. 6(b), there is no evidence of the microbarom signal in the first separated signal. However, when comparing the top and bottom graphs in Fig. 6(d) we see that there is evidence of a microbarom signal *buried* in the second

ICA separated signal. Specifically, in the 0.1 to 0.2 Hz region there are two spectral peaks in both PSDs. Probably more profound is the observation made when comparing the bottom graph in Fig. 5(b) for the first separated signal, to the top graph in Fig. 6(c) for the second separated signal. Upon first glance when comparing these two spectra, they appear to be almost the same. However, when they are rescaled, shown in the top graph in Fig. 6(b) for the first separated signal, and the top graph in Fig. 6(d) for the second separated signal, the differences in the spectra can be seen. In fact, the correlation coefficient computed between the two time-domain signals is very low, specifically, 1.7992×10^{-4}. But more importantly is the overlapping spectra that exists between the two signals in the frequency range from 0.01 Hz to 0.02 Hz. In spite of this spectral overlap, ICA can separate the signals.

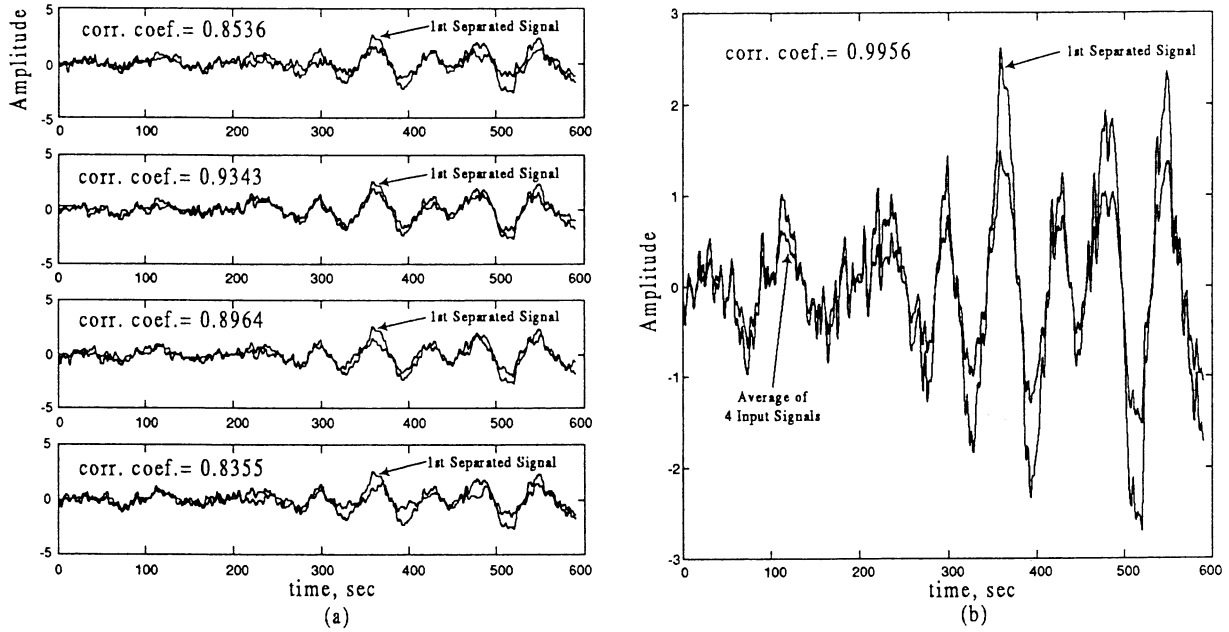

Figure 4. (a) First ICA separated signal superimposed on the four input signals. (b) First ICA separated signal superimposed on the average of the four input signals.

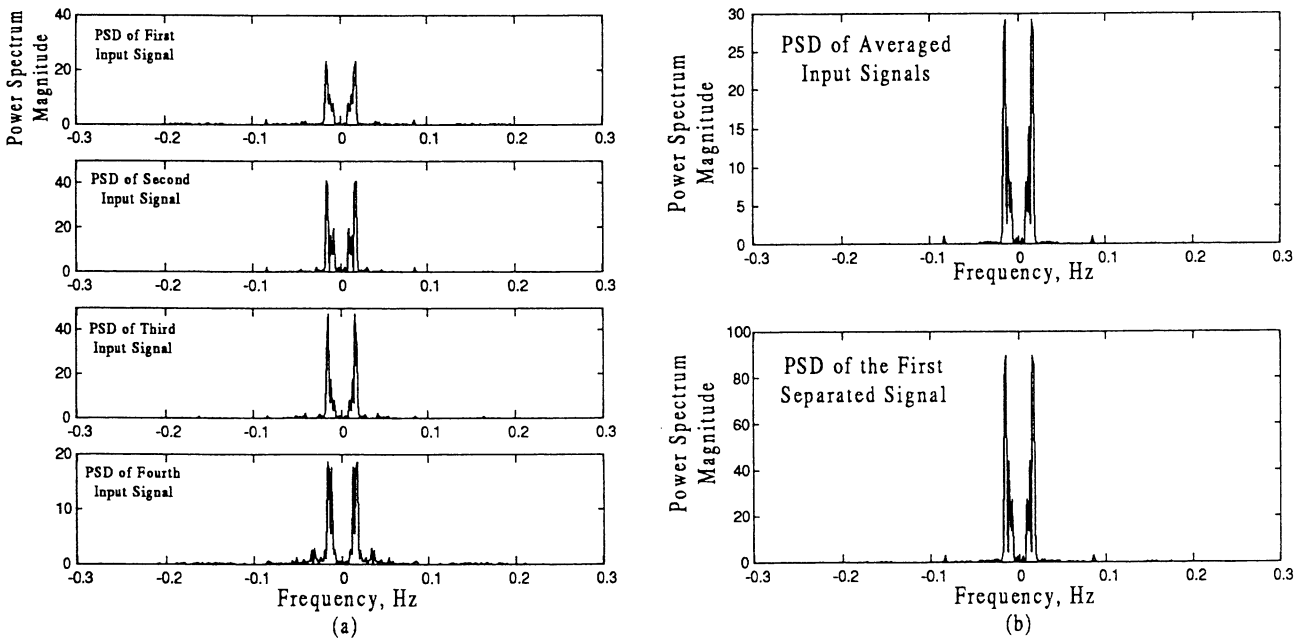

Figure 5. (a) PSD of the four ICA input signals. (b) Top Graph: PSD of the average of the four ICA input signals, and Bottom Graph: PSD of the first ICA separated signal.

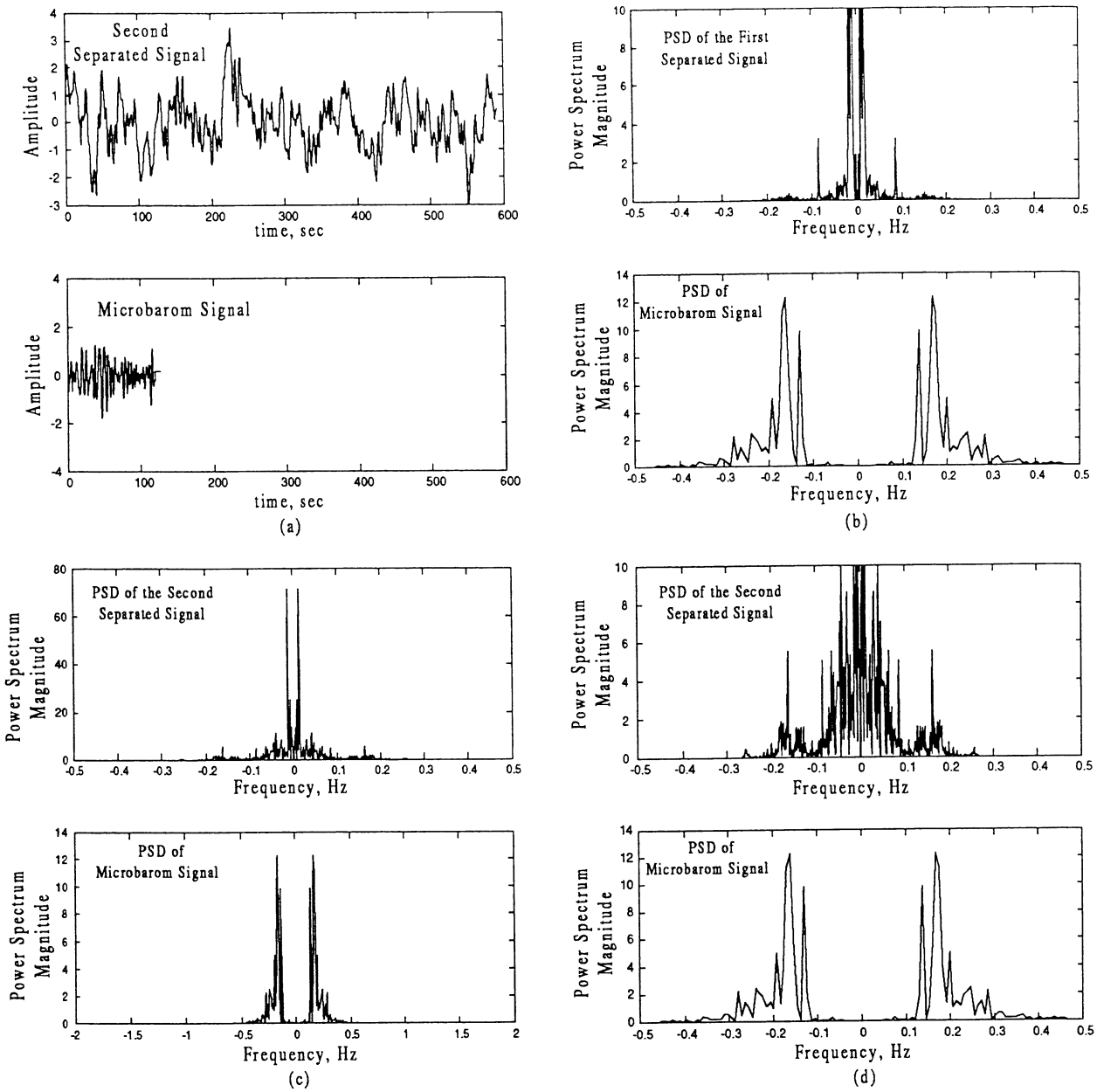

Figure 6. (a) Top Graph: Second ICA separated signal, Bottom Graph: microbarom infrasound signal recorded on the T-array. (b) Top Graph: PSD of the first separated signal, and Bottom Graph: PSD of the microbarom signal. (c) Top Graph: PSD of the second separated signal, and Bottom Graph: PSD of the microbarom signal. (d) Top Graph: Rescaled PSD of the first separated signal, and Bottom Graph: Rescaled PSD of the microbarom signal.

CONCLUSIONS AND RECOMMENDATIONS

We have shown the feasibility of applying Independent Component Analysis (ICA), using a neural learning algorithm, to single station infrasound signals for the purpose of separating signal from noise. The second example presented showed that infrasound signals recorded on a 4-sensor array (F-array) could be separated into a signal of interest (volcano event) and background noise. The signal of interest was shown to be the volcano infrasound signal. It was also discovered that the background noise contained a microbarom component. This was determined by observing the noise signal in the frequency domain (i.e., its PSD) compared to the PSD of a typical microbarom signal. It was also observed that the two separated signals had overlapping spectra. In spite of this,

the ICA was able to separate the signals. Further research in this area will involve investigating other infrasound recordings from the historical database [19, 20] for other types of events. When enough separated signals are generated for different infrasound events, this data set will be used to train and test a neural network classifier.

REFERENCES

[1] C. Jutten and J. Herault, "Independent component analysis (INCA) versus principal components analysis," in *Signal Processing* IV: *Theories and Applications*, Proc. *EUSIPCO*-88, J. Lacoume, *et al.*, Eds., Amsterdam, The Netherlands, Elsevier, pp. 643-646, 1988.

[2] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetric architecture," *Signal Processing*, **24**, pp. 1-10, 1991.

[3] P. Comon, "Independent component analysis," in Proceedings of the *International Signal Processing Workshop on Higher-Order Statistics,* Chamrousse, France, pp. 111-120, 1991. [Republished in *Higher-Order Statistics*, J.L. Lacoume, Ed., Amsterdam, The Netherlands, Elsevier, pp. 29-38, 1992.]

[4] P. Comon, "Independent component analysis-A new concept?," *Signal Processing*, **36**, pp. 287-314, 1994.

[5] F.M. Ham, T.A. Leeney, H.M. Canady, and J.C. Wheeler, "Discrimination of Volcano Activity Using Infrasonic Data and a Backpropagation Neural Network," In Proceedings of the *SPIE Conf. on Appl. and Sci. of Compt. Intel. II*, Eds. K.L Priddy, P.E. Keller, D.B. Fogel, and J.C. Bezdek, Orlando, FL, **3722**, pp. 344-356, 1999.

[6] F.M. Ham, T.A. Leeney, H.M. Canady, and J.C. Wheeler, "A Radial Basis Function Neural Network for Classification of Infrasonic Events for the CTBT IMS," *American Geophysical Union Conference, EOS Transactions, AGU*, vol. 80, no.17, Boston, MA, June 1-4, (Abstract A52A-01), p. S71, 1999.

[7] I.T. Jolliffe, *Principal Component Analysis*, New York: Springer-Verlag, 1986.

[8] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: Theoretical results and applications," *Proc. of the IEEE*, **79**, pp. 278-305, 1991.

[9] C.L. Nikias and J.M. Mendel, "Signal processing with higher-order spectra," *IEEE Signal Processing Mag.*, **10**, pp. 10- 37, 1993.

[10] J. Karhunen, E. Oja, L. Wang, R. Vigario, and J. Joutsensalo, "A class of neural networks for independent component analysis," *IEEE Trans. on Neural Networks*, **8**, pp. 486-504, 1997.

[11] J.-F. Carduso and B.H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Signal Processing*, **44**, pp. 3017-3030, 1996.

[12] J. Karhunen, "Neural approaches to independent component analysis and source separation," In *Proc. 4th European Symp. Artificial Neural Networks, ESANN'96*, Bruges, Belgium, pp. 249-266, April 1996.

[13] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, **10**, pp. 626-634, 1999.

[14] E. Oja, H. Ogawa, and J. Wangviwattana, "Learning in nonlinear constrained networks," in *Artificial Neural Networks Proc. ICANN-91*, T. Kohonen, *et al.* Eds., Amsterdam, The Netherlands, North-Holland, pp. 385-390, 1991.

[15] J. Karhunen and J. Joutsensalo, "Representation and separation of signals using nonlinear PCA type learning," *Neural Networks*, **7**, pp. 113-127, 1994.

[16] J. Karhunen and J. Joutsensalo, "Generalizations of principal component analysis, optimization problems, and neural networks," *Neural Networks*, **8**, pp. 549-562, 1995.

[17] L. Wang and E. Oja, "A unified neural bigradient algorithm for robust PCA and MCA," *Int. J. Neural Syst.*, **7**, pp. 53-67, 1996.

[18] L. Wang, J. Karhunen, and E. Oja, "A bigradient optimization approach for robust PCA, MCA, and source separation," In *Proc. 1995 IEEE Int. Conf. on Neural Networks,* Perth, Australia, pp. 1684-1689, Nov. 1995.

[19] C.R. Wilson, J.V. Olson, and R. Richards, "Library of typical infrasonic signals," Report Prepared for ENSCO (subcontract no. 269343-2360.009) **1-4**, 1996.

[20] K.D. Hutchenson, "Acquisition of historical infrasonic data," Final Technical Report, Contract No. F08650-95-D-0033, ARS-97-012, ENSCO, Melbourne, FL, 1997.