Blind Separation of Multiple Vehicle Signatures In Frequency Domain

M. R. Azimi-Sadjadi *and S. Srinivasan

Information System Technologies, Inc, Fort Collins, CO 80521.

ABSTRACT

This paper considers the problem of classifying ground vehicles using their acoustic signatures recorded by unattended passive acoustic sensors. Using these sensors, acoustic signatures of a wide variety of sources such as trucks, tanks, personnel, and airborne targets can be recorded. Additionally, interference sources such as wind noise and ambient noise are typically present. The proposed approach in this paper relies on the blind source separation of the recorded signatures of various sources. Two different frequency domain source separation methods have been employed to separate the vehicle signatures that overlap both spectrally and temporally. These methods rely on the frequency domain extension of the independent component analysis (ICA) method and a joint diagonalization of the time varying spectra. Spectral and temporal-dependent features are then extracted from the separated sources using a new feature extraction method and subsequently used for target classification using a three-layer neural network. The performance of the developed algorithms are demonstrated on a subset of a real acoustic signature database acquired from the US Army TACOM-ARDEC, Picatinny Arsenal, NJ.

Keywords: Blind Source Separation, Frequency Domain Independent Component Analysis, Target Classification, Acoustic Sensors

1. INTRODUCTION

Unattended passive acoustic sensors[1-3] are among the widely used sensors for remote battlefield surveillance, situation awareness and monitoring applications. Together with dedicated digital signal processing (DSP), these small and cost effective sensors can provide real-time information about different types of ground and airborne targets. Generally, there can be a wide variety of target types in battlefield depending on the specific mission, e.g. ground targets (trucks, tanks, etc), airborne targets (helicopters, missiles, airplanes), or personnel in urban areas. These sources typically have signatures that overlap both temporally and spectrally.

The primary goal of this study is to successfully separate the acoustic signatures of multiple sources not only to determine the number of sources but also be able to extract tonal and frequency dependent features for subsequent classification. Two different frequency domain blind source separation (BSS) algorithms[4-7] are studied and extended for this problem. In these methods, the recorded microphone signals are first transformed to the frequency domain using short time Fourier transform (STFT). This converts a convolutive mixture (with unknown fractional time delays) in the time domain to a complex linear mixture in the frequency domain, which can then be used in conjunction with BSS at each frequency bin to separate the sources. The first method[4] uses frequency domain independent component analysis (ICA) at each frequency bin independently. To account for the scaling at different frequencies, which can result in distortion of the spectrum at the output, the separated outputs are filtered by the pseudo-inverse of the separation filter. Moreover, to account for the permutation at different frequency bins, which may lead to remixing of the signals in the final output, the spectral envelope at different frequency bins are correlated to associate the separated sources at adjacent frequencies within a subband. In the second method[5-,7] a joint diagonalization BSS method was proposed that exploits second-order statistics of nonstationary observations. Unlike most of the BSS algorithms, this algorithm takes advantage of the inherent nonstationarity of the sources to achieve blind separation. Since the focus of the algorithm is on nonstationarity, certain assumptions are made to simplify the developments. In particular, it is assumed that the sources are Gaussian and temporally independent.

*mo@infsyst.biz; Phone: 970 224 2556; Fax 970 224 2556

Indeed for this particular application, the nonstationarity is more relevant than the non-Gaussianity due to temporal and spectral variations of the signatures as sources move in the field. This fact will be demonstrated in our test results in this paper. Furthermore, a benchmarking with the results of the frequency domain ICA-based BSS in[4] is also provided to select the best algorithm for subsequent source classification. A multi-layer back-propagation network (BPNN) is employed as a classifier. A set of spectral and temporal features is extracted from each separated source and then applied to the BPNN classifier. This network classifies the sources into one of four possible classes, namely heavy tracked, heavy wheeled, light tracked and light wheeled. Test results on real acoustic signature data sets acquired from US Army TACOM-ARDEC are presented to show the effectiveness of the proposed schemes.

2. FREQUENCY-DOMAIN ICA-BASED METHOD

This section presents a review of the BSS based on frequency domain ICA method[4] and our modifications to some of the steps in the algorithm. In this method, the recorded microphone signals are first transformed to the frequency domain using STFT. This converts a convolutive mixture with unknown number of taps and delays to a complex linear mixture at each frequency bin. Once transformed in the frequency domain, a subspace whitening process is applied not only to reduce the effects of reflections and ambient noise but also to sphere the data.[4] Next, the instantaneous complex ICA is applied to the output of the subspace filter. In the usual instantaneous ICA,[8] arbitrary permutation and scaling of the output is allowed. However, in the frequency domain procedure, different permutation at different frequency bins leads to remixing of the signals in the final output. Moreover, different scaling at different frequencies can result in distortion of the spectrum at the output. Thus, scaling and permutation are extremely important in the frequency domain case. For the scaling, the method in[11] is used where the separated output is filtered by the inverse of the separation filter. For the permutation, spectral envelope at different frequency bins are correlated to associate the separated sources at adjacent frequencies. The overview of the entire process system and the constituent subsystems are shown in Figure 1. These subsystems and their operations are briefly described in the following sections.

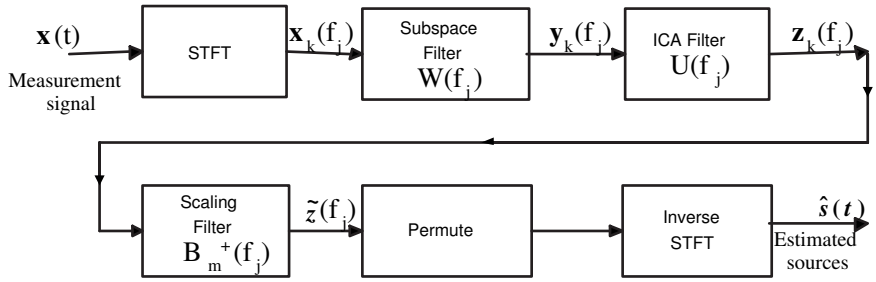


Figure 1. The BSS system

2.1. STFT and Signal Model in Frequency Domain

Let us consider an array of M sensors that receive the wavefield generated by d wideband sources in the presence of an arbitrary noise. The array geometry can be arbitrary but known to the processor. The sources $s_1(t), s_2(t), \ldots, s_d(t)$ are assumed to be statistically independent. The noise is also assumed to be independent of the source signals.

The array output vector $\mathbf{x}(t)$ is first decomposed into narrowband frequency components by taking the short time frequency transform (STFT) of each recorded signal and treating each frequency bin independently. That is, the array output $\mathbf{x}(t)$, is partitioned into K overlapping windows (e.g. Gaussian or Hamming) of duration T and overlap ΔT seconds and then DFT of the windowed signal is taken. For the j^{th} frequency component, f_j, at the k^{th} window, the signal $\mathbf{x}_k(f_j)$ can be represented by

$$\mathbf{x}_k(f_j) = \mathrm{A}(f_j)\mathbf{s}_k(f_j) + \mathbf{n}_k(f_j) \tag{1}$$

where $\mathrm{A}(f_j)$ is the $M \times d$ mixing matrix, $\mathbf{s}_k(f_j)$ and $\mathbf{n}_k(f_j)$ vectors consist of the source and noise spectra. Note that the (m,n) element of $\mathrm{A}(f_j)$ matrix represents the transfer function from the nth source to the mth microphone in the array. It is assumed that $M > d$ and that the rank of $\mathrm{A}(f_j)$ is equal to d for any frequency. As can be seen, the convolutive mixture with unknown propagation delays between sources and microphones is now converted to a linear complex-valued mixture in (1), which is more amenable to ICA processing.

2.2. Subspace Filter

To reduce the effect of noise (assume $M > d$) and at the same time sphere the data, we apply a subspace-based filter to $\mathbf{x}_k(f_j)$ to reduce the effect of noise and whiten the input signal. First, the spatial covariance matrix for the jth narrowband component $\mathbf{x}_k(f_j)$ is computed as

$$\mathrm{R}_x(f_j) = \mathrm{A}(f_j)\mathrm{R}_s(f_j)\mathrm{A}^H(f_j) + \mathrm{R}_n(f_j), \qquad j = 1, 2, \ldots, J. \tag{2}$$

where $\mathrm{R}_s(f_j)$ and $\mathrm{R}_n(f_j)$ are spatial covariance matrices of the source and noise, respectively. The noise and sources are assumed to be independent and further noise $\mathbf{n}$ is assumed to be spatially white. Then, the eigenvalue decomposition of $\mathrm{R}_x(f_j)$ is performed using $\mathrm{R}_x(f_j) = \mathrm{E}\Lambda\mathrm{E}^{-1}$ where E is the eigenvector matrix and $\Lambda = diag[\lambda_1, \cdots, \lambda_M]$ with λ_i being the ith eigenvalue of matrix $\mathrm{R}_x(f_j)$. Now, if E_s is the corresponding signal subspace matrix with $Span\{\mathrm{A}\} = Span\{\mathrm{E}_s\}$, then the subspace filtering process yields,

$$\mathbf{y}_k(f_j) = \mathrm{W}(f_j)\mathbf{x}_k(f_j) \tag{3}$$

where $\mathrm{W}(f_j) = \Lambda_s^{-1/2}\mathrm{E}_s^H(f_j)$ and $\Lambda_s = diag[\lambda_1, \cdots, \lambda_d]$ is the matrix with d dominant (signal) eigenvalues. It is easy to see that the correlation matrix of $\mathbf{y}_k(f_j)$ is identity, i.e.

$$\begin{aligned} E[\mathbf{y}_k(f_j)\mathbf{y}_k^H(f_j)] &= \Lambda_s^{-1/2}\mathrm{E}_s^H E[\mathbf{x}_k(f_j)\mathbf{x}_k^H(f_j)]\mathrm{E}_s\Lambda_s^{-H/2} \\ &= \Lambda_s^{-1/2}\mathrm{E}_s^H E\Lambda E^{-1}\mathrm{E}_s\Lambda_s^{-H/2} \\ &= I_d \end{aligned} \tag{4}$$

which shows that the subspace method is indeed a whitening procedure. Meanwhile[4] proved that the subspace method also acts as a self-organizing beamformer that reduces the energy of the noise subspace.

2.3. Frequency Domain ICA and Learning Rule

The instantaneous complex-valued ICA is applied to the output of the subspace filter $\mathbf{y}_k(f_j)$ to obtain the separated sources. The frequency domain ICA algorithm is briefly described below. If we denote the demixing matrix at frequency f_j by $U(f_j)$, then the output $\mathbf{z}_k(f_j)$ of the frequency domain ICA algorithm is,

$$\mathbf{z}_k(f_j) = U(f_j)\mathbf{y}_k(f_j) \tag{5}$$

$$= U(f_j)W(f_j)\mathbf{x}_k(f_j) \tag{6}$$

$$= B(f_j)\mathbf{x}_k(f_j) \tag{7}$$

where $B(f_j) := U(f_j)W(f_j)$ is referred to as the separation filter. The learning rule for the ICA filter $U(f_j)$ is based upon the extension of the Amari's rule,[8].[9] This is given by,

$$U_{k+1}(f_j) = U_k(f_j) + \eta[I - \phi(\mathbf{z}_k(f_j)\mathbf{z}_k^H(f_j)]U_k(f_j), \tag{8}$$

where η is the learning rate, and the score function $\phi(\mathbf{z})$ is defined[4] as

$$\phi(\mathbf{z}) = 2\tanh(G \cdot \mathrm{Re}(\mathbf{z})) + 2j\tanh(G \cdot \mathrm{Im}(\mathbf{z})), \tag{9}$$

where G is the gain for the nonlinear score function. Clearly, any permuted and/or scaled version of the separation filter obtained by the frequency domain ICA is an acceptable solution for the BSS. Thus, the permutation and scaling problems should carefully been addressed.

2.4. Scaling Problem

It was proposed that the scaling problem could be solved by filtering individual output of the separation filter by the inverse of $B(f_j)$, which is the estimate of the mixing matrix. In this report, the pseudo-inverse of $B(f_j)$, denoted as $B^{\dagger}(f_j)$ is used instead of the actual inverse. The contribution of the n^{th} source in the recorded signal at the m^{th} microphone is,

$$\tilde{z}_{k,m,n}(f_j) = B^{\dagger}_{m,n}(f_j) z_{k,n}(f_j) \tag{10}$$

where $B^{\dagger}_{m,n}(f_j)$ denotes the $(m,n)^{th}$ element of matrix $B^{\dagger}(f_j)$ and $z_{k,n}(f_j)$ represents the n^{th} separated source signal. Hence the contributions from all d sources at the m^{th} microphone could be defined by,

$$\tilde{\mathbf{z}}_k, m(f_j) = \begin{pmatrix} B^{\dagger}_{m,1}(f_j) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & B^{\dagger}_{m,d}(f_j) \end{pmatrix} \begin{pmatrix} z_{k,1}(f_j) \\ \vdots \\ z_{k,d}(f_j) \end{pmatrix} \tag{11}$$

$$= B^{\dagger}_{m}(f_j)\mathbf{z}_k(f_j) \tag{12}$$

where $B^{\dagger}_{m}(f_j) = diag[B^{\dagger}_{m,1}(f_j), \cdots, B^{\dagger}_{m,d}(f_j)]$ is a $d \times d$ diagonal matrix. Note that we have $x_{k,m}(f_j) = \sum_{i=1}^{d} \tilde{z}_{k,m,i}(f_j)$, which is the recorded signal (frequency bin f_j) at the m^{th} microphone.

2.5. Permutation Problem

Let us denote the estimate of the mixing matrix obtained using the pseudo-inverse $\mathrm{B}^{\dagger}(f_j)$ as $\hat{\mathrm{A}}(f_j) = \mathrm{B}^{\dagger}(f_j)$. Then define the estimate of the mixing matrix multiplied by the arbitrary permutation matrix P by,

$$\mathrm{C}^{T}(f_j) = P\mathrm{A}^{T}(f_j) \tag{13}$$

The permutation $P\mathrm{A}^{T}(f_j)$ exchanges the rows of $\mathrm{A}^{T}(f_j)$. If the column vectors of $C(f_j)$ denoted as $C(f_j) = [\mathbf{c}_1(f_j), \cdots, \mathbf{c}_d(f_j)]$. The cosine of the angle α_n between two vectors, $\mathbf{c}_n(f_j)$ and $\mathbf{c}_n(f_{j-1})$ at two adjacent frequency bins f_j and f_{j-1} is defined as,

$$cos(\alpha_n) = \frac{\mathbf{c}^{H}_{n}(f_j)\mathbf{c}_n(f_{j-1})}{\|\mathbf{c}_n(f_j)\|.\|\mathbf{c}_n(f_{j-1})\|} \tag{14}$$

A permutation matrix P is then chosen that maximizes the cost function,

$$F(P) = \frac{1}{d}\sum_{n=1}^{d} cos(\alpha_n) \tag{15}$$

It must be mentioned that this procedure assumes that the mixing matrices at several adjacent frequencies remain unchanged. However, at some frequencies, this assumption may not hold if ICA failed. Since the permutation at frequency f_j is determined based on only the information the adjacent frequencies f_j and f_{j-1}, and the permutation is solved iteratively with increasing frequency, once the permutation at certain frequency fails, the permutation in the succeeding frequencies may also fail. Hence to prevent this, the cost function in (15) is calculated at several adjacent frequencies $f_{j-i}, i \in [1, K]$. Let us denote the value of the cost function evaluated between f_j and f_{j-i} as $F_i(P)$. If the largest value of the cost function at some P is close to the cost function at other $P's$ then the permutation matrix estimated may not be reliable. Based on this, the following confidence measure is defined,

$$C(i) = \max_{P \in\ \Omega} F_i(P) - \max_{P \in\ \Omega'} F_i(P) \tag{16}$$

where Ω denotes the set of all possible permutation matrices P and Ω' denotes Ω excluded with $arg\max_{P \in\ \Omega} F_i(P)$. If $\hat{i} = \max_i C(i)$, then the appropriate permutation matrix is,

$$C(i) = arg\max_{P} F_{\hat{i}}(P) \tag{17}$$

2.6. Final Filtering Matrix

Using the matrices obtained above, the final filtering matrix, of size $D \times D$, in the frequency domain can be written as,

$$F(f_j) = P(f_j)B_m^+(f_j)B(f_j) \tag{18}$$

and the separated source at frequency f_j is,

$$\mathbf{z}_k(f_j) = F(f_j)X_k(f_j) \tag{19}$$

2.7. Time Domain Reconstruction of Separated Source

The equation (19) is to be applied for all frequencies. In our approach, the separated sources $\hat{\mathbf{s}}(t)$ are reconstructed in time domain[12] using,

$$\hat{\mathbf{s}}(t) = \frac{1}{N}\sum_{k=0}^{N-1} w_k(t)\sum_{j=1}^{J}\mathbf{z}_k(f_j)exp(\frac{i2\pi f_j t}{N}) \tag{20}$$

where J is the number of frequency bins and N is the number of time windows. The inner summation represents the inverse DFT of $\mathbf{z}_k(f_j)$.

3. BSS BASED ON NONSTATIONARITY OF SOURCES

In this section, we present a blind separation[5-7] of convolutive mixtures based on the joint diagonalization of time varying spectral matrices of the observations. The convolutive mixture in the time domain is first represented as a linear mixture in the frequency domain as in (1). The objective in blind separation is to separate signals from the original mixtures such that the separated ones are as mutually independent as possible. In this algorithm, by adopting a second order approach, the focus is maintained only on the inter-correlations between the reconstructed sources at all lags or equivalently the inter-spectra between them at all frequencies. However, as this approach assumes the nonstationary of the signals, the time varying spectra, i.e. the localized spectra around each given time point are considered.

3.1. Objective Function

Let the observation sequence be divided into different blocks of data and $S_{\mathbf{x}}(k, f_j)$ be the spectra of the k^{th} block at frequency bin f_j. To perform source separation, the natural idea is to find matrix $G(f_j)$ at each frequency f_j such that $G(f_j)\hat{S}_{\mathbf{x}}(k, f_j)G^*(f_j)$ is as close as possible to diagonal, at different time points k, where $\hat{S}_{\mathbf{x}}(k, f_j)$ is the estimate of $S_{\mathbf{x}}(k, f_j)$. For the fixed matrix $G(f_j)$, the diagonality measure is given by,

$$\frac{1}{2}\{log\ det\ diag[G(f_j)\hat{S}_{\mathbf{x}}(k, f_j)G^*(f_j)] - log\ det[G(f_j)\hat{S}_{\mathbf{x}}(k, f_j)G^*(f_j)]\} \tag{21}$$

The second term can be written as $2log|detG(f_j)| + log\ det\hat{S}_{\mathbf{x}}(k, f_j)$ and the term $log\ det\hat{S}_{\mathbf{x}}(k, f_j)$ is unchanged for any choice of $G(f_j)$ and can be dropped. Therefore, the diagonality criterion taking into account different blocks of data is,

$$\sum_k(\frac{1}{2}log\ det\ diag[G(f_j)\hat{S}_{\mathbf{x}}(k, f_j)G^*(f_j)] - log|detG(f_j)|) \tag{22}$$

This criterion is to be minimized with respect to $G(f_j)$ to obtain the final separation matrix at each frequencies f_j. Algorithms for the joint diagonalization of several positive matrices is used to solve for the above problem.

3.2. Spectral Estimation

The signal in the time domain is first divided into different overlapping blocks of data of size N. The periodogram of each block is then computed from the DFT of the data blocks. The periodogram of the $(k+1)^{th}$ data block of size N is,

$$P_{\mathbf{x}}(k+1, f_j) = \frac{1}{N}\left[\sum_{l=kN+1}^{kN+N} \mathbf{x}(l)e^{2\pi j f_j l}\right]\left[\sum_{l=kN+1}^{kN+N} \mathbf{x}(l)e^{2\pi j f_j l}\right]^* \tag{23}$$

The frequencies are taken as $f_j = \frac{j}{N},\ j \in [0, N/2]$. The time varying spectrum at the mid point $l_k = kN - (N-1)/2$ of the k^{th} block is then estimated as,

$$\hat{S}_{\mathbf{x}}(l_k, f_j) = \frac{1}{2m+1}\sum_{i=-m}^{m} P_{\mathbf{x}}(k, f_{j-i}) \tag{24}$$

where $m \leq N/2 - 1$ is a bandwidth parameter.

3.3. Algorithm for Joint Diagonalization

The algorithm for joint approximate diagonalization[5] of several positive matrices computes a matrix $G(f_j)$ that minimizes the criterion in (22). The algorithm uses the classic Jacobi approach of operating by successive transformations on each pair of rows $G(f_j)$, but the transformations are not constrained to be orthogonal. Explicitly let $G(f_j)_i$ and $G(f_j)_l$ be a pair of rows of $G(f_j)$, the algorithm changes $G(f_j)$ into a new matrix with these rows given by,

$$\begin{bmatrix} G(f_j)_i \\ G(f_j)_l \end{bmatrix} \leftarrow \begin{bmatrix} G(f_j)_i \\ G(f_j)_l \end{bmatrix} - T_{il}\begin{bmatrix} G(f_j)_i \\ G(f_j)_l \end{bmatrix} \tag{25}$$

where the other rows are unchanged. Here, T_{il} is a 2×2 matrix that can be chosen such that the criterion is sufficiently decreased. The procedure is then repeated with another pair of rows. The transformation matrix is computed as follow. Define the quantities,

$$g_{il} = \sum_k \nu_k \frac{(G(f_j)\hat{S}_{\mathbf{x}_k}(f_j)G(f_j)^T)_{il}}{(G(f_j)\hat{S}_{\mathbf{x}_k}(f_j)G(f_j)^T)_{ii}},\ w_{il} = \sum_k \nu_k \frac{(G(f_j)\hat{S}_{\mathbf{x}_k}(f_j)G(f_j)^T)_{ll}}{(G(f_j)\hat{S}_{\mathbf{x}_k}(f_j)G(f_j)^T)_{ii}}$$

$$\begin{bmatrix} h_{il} \\ h_{li} \end{bmatrix} = \begin{bmatrix} w_{il} & 1 \\ 1 & w_{li} \end{bmatrix}^{-1}\begin{bmatrix} g_{il} \\ g_{li} \end{bmatrix} \tag{26}$$

where $\sum_k \nu_k = 1$.

$$T_{il} = \frac{2}{1+\sqrt{1-4h_{il}h_{li}}}\begin{bmatrix} 0 & h_{il} \\ h_{li} & 0 \end{bmatrix} \tag{27}$$

Note that the transformation (25) generated by (27) always leads to the decrease of the objective function unless $g_{il} = g_{li} = 0$, which corresponds to the $(i,l)^{th}$ of the gradient of the objective function.

3.4. Solving the Permutation Ambiguity

The procedure employed to solve the permutation ambiguity is same as the one described for the frequency domain ICA in Section 2.1.5. Final filtering is conducted in the time domain to avoid time domain aliasing. The time domain filters are obtained as the IDFT of $G(f_j)$ as,

$$g_{mn}(l) = IDFT\{G_{mn}(f_j)\}_{j=0}^{N/2} \tag{28}$$

where $G_{mn}(f_j)$ and $g_{mn}(l)$ denote the $(m,n)^{th}$ element of the frequency domain and time domain filters, respectively. Once the filter responses $g_{mn}(l)$'s are determined, the separated sources are obtained by convolving them with the recorded signals.

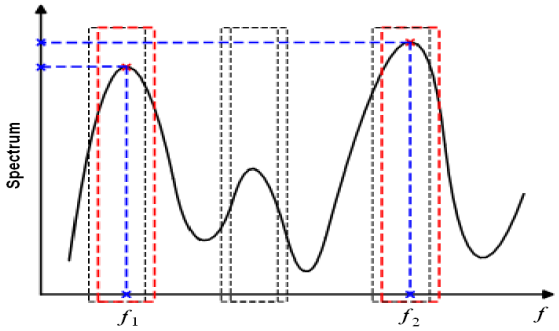


Figure 2. Peak finding using a sliding window. The detected peaks are located at frequencies f_1 and f_2.

4. SUBBAND FEATURE EXTRACTION

Observation of the spectrograms of the signatures of the ground targets of interest reveal the fact their frequency behavior is not truly wideband and that target indications exist in disjoint narrow frequency subbands. That is, the spectrum of the recorded signal at each time instant has disjoint peaks at frequencies where target indications are present. On the other hand, wind interference has a relatively wideband frequency behavior which is somewhat exponentially decaying as frequency increases. Consequently, the presence of ground target may be detected by applying a peak finding algorithm to the spectrum of a signal recorded by a microphone at each time instant. The main idea behind our proposed subband feature extraction approach is to detect the frequency peaks and the subbands that contain potential target information and then extract time and frequency dependent features from these detected subbands. The steps of the developed feature extraction algorithm may be summarized as follows:

1. The observation interval of 10 seconds is divided into blocks of overlapping segments of 128 samples with 50% overlap. Note that observation interval of 10 seconds long seems to be optimum as the decision about the class membership cannot be made in smaller time interval and enough clues need to be gathered before final decision making.
2. STFT is performed in each of the data blocks and up to $N = 5$ frequency peaks, in the frequency range of 1-256 Hz, that are above a certain threshold are selected in each block. The frequency separation is 1 Hz.
3. Record the detected frequency peaks in the time-windowed spectra.

In Step 2, to detect the peaks of the time-windowed spectra, a peak finding process is used (see Figure 2). This scheme uses a sliding window (shown in dashed-line rectangle) to localize the maximum value of the time-frequency spectrum within the window. If the maximum happens at the center of the window and its value is greater than a 'threshold' the frequency that corresponds to this maximum is recorded as a peak point or a detection frequency. The window is then moved by 1 Hz and the procedure is repeated. The windows specified in red are those for which peaks at frequencies f_1 and f_2 are detected.

A simple way to determine a threshold for peak finding, which is somewhat adapted to the noise level in each sliding window procedure, is to find the median value of the portion of the spectra which lies inside the sliding window and then use a fixed pre-specified percentage (120%) of the median as the local threshold. We refer to this as "median-based thresholding". A potential peak is selected as a frequency peak if it is greater than this local threshold.

In our algorithm, the peak finding is employed for a time-windowed spectrum which represents the time-frequency behavior of the acoustic signatures over a very short period of time (1/8th of a second). In order to determine temporal variations or consistency of the detected frequencies over the observation period, a simple

histogram-based scheme is devised where the number of bins in the histogram (32) corresponds to the number of 8Hz subbands. Since the acoustic signatures of targets of interest don't have useful spectral information below 17Hz the first two subbands are not used. This yields a 30-dimensional feature vector where the ith element represents the number of occurrence of frequency peaks that lie within the ith subband, i.e. frequency range $(8(i-1)+1, 8i)$Hz. The overall pattern of these elements is representative of temporal variations or consistency of the subband spectral features over the observation period. This feature vector is then applied to the classifier.

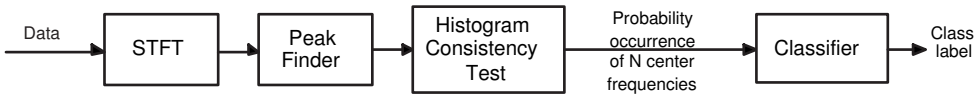


Figure 3. Procedure for feature extraction and classification.

5. TEST RESULTS

5.1. Results on Source Separation

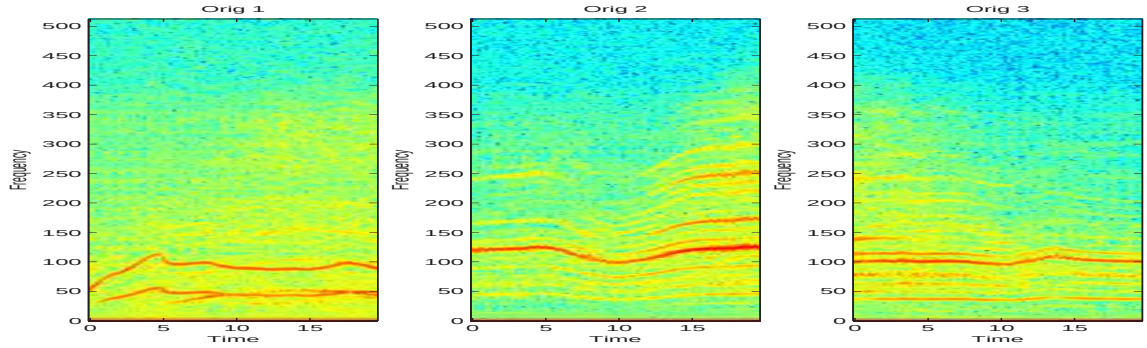
The data set for this study is a subset of an acoustic signature database acquired from the US Army TACOM-ARDEC. The data has been collected using three wagon wheel-type pattern array of five identical elements (microphones) with 2ft radius at a sampling rate of $f_s = 1024$Hz. In the following subsections, two studies are conducted to compare the performance of the two source separation algorithms presented in this paper and show their promise and shortfalls for vehicle classification.

5.1.1. Synthesized Cases

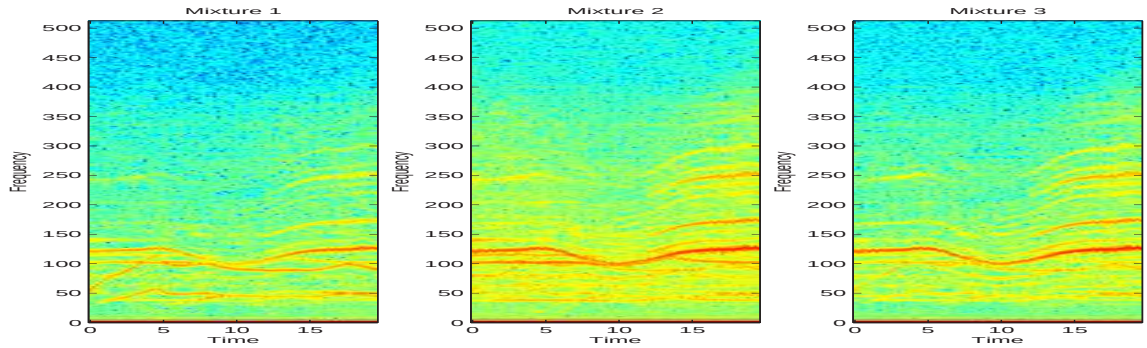
First, we benchmark the performance of the nonstationary BSS algorithm against the frequency domain ICA-based method on synthesized data. Tables 1 and 2 give different parameters for implementing the two algorithms. Three known source signatures extracted from three real single-target runs in the target signature database were convolutively mixed to form synthetic observations. Source separation based on the two BSS methods was then performed on the synthetically generated data. Let $s_1(t)$, $s_2(t)$ and $s_3(t)$ be the three known source signals collected by the center microphone of a 5-element array in the three single-target runs. Then, the observation vector is generated using the source mixing model,

$$\mathbf{x(t)} = \begin{bmatrix} 6s_1(t) + 3s_2(t) + 2s_3(t) \\ 3s_1(t+1) + 7s_2(t+1) + 4.5s_3(t+1) \\ 3s_1(t+2) + 7s_2(t+2) + 1.5s_3(t+2) \end{bmatrix} \tag{29}$$

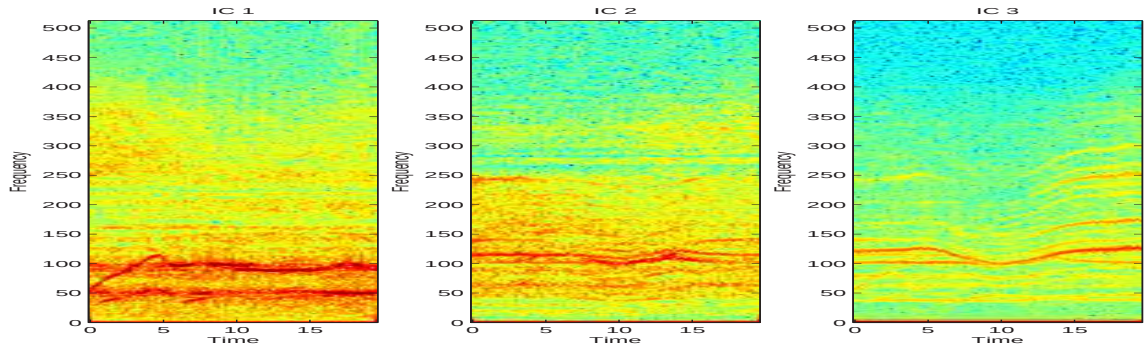
Figures 4 (a) and (b) show the spectrograms of the known sources and the synthetically generated convolutive mixtures, respectively. The spectrograms of the three separated sources using the frequency domain ICA method and nonstationary BSS method are shown in Figures 4 (c) and (d), respectively. As can be observed from Figure 4 (c), the signatures of the separated sources using the frequency domain ICA method share some common spectral features of all the three original sources. This implies that the source separation using this method has failed even in this simple case. However, from Figure 4 (d), it can be seen that the spectrograms of the separated sources closely resemble those of the original sources with no visible common features among them. The results clearly indicate that BSS based on the nonstationarity assumption works much better than that based on the frequency domain ICA. There could be many reasons for imperfect performance of the frequency domain ICA method including but not limited to: (a) possible failure of the ICA algorithm at certain frequency bins and its effects on determining permutation matrix and performing source association, (b) inappropriate and erroneous choice of the pseudo-inverse of $B(f_j)$ as the estimate of the mixing matrix which not only ignores the high order statistical properties of the data but also does not support the structure of the array steering matrix, and (c) the stationarity assumption.



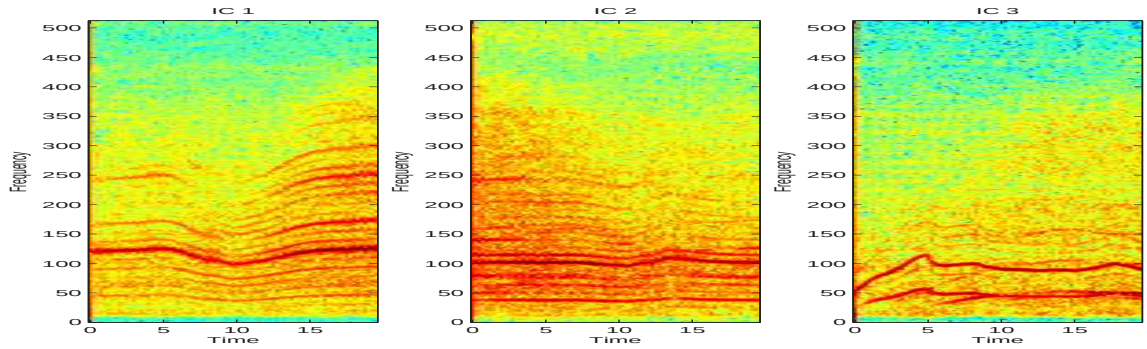
(a) Known source signals



(b) Convolutively mixed observations.



(c) Sources separated using frequency domain ICA-based method.



(d) Sources separated using non-stationary BSS method.

Figure 4. Source separation on synthetically generated convolutive mixture.

Table 1. Parameters of the frequency domain ICA-based method

Sampling Frequency	1024 Hz
Length of STFT	512
Shift of STFT	256
Window function	Hamming
Learning rate, η	0.00001
Gain for score function, G	100
Reference range in permutation, K	5
Number of sensors, M	5
number of sources, d	2

Table 2. Parameters of the non-stationary based method

block size, N	512
overlap	256
bandwidth	3
Reference range in permutation, K	5
Number of sensors, M	5
number of sources, d	2

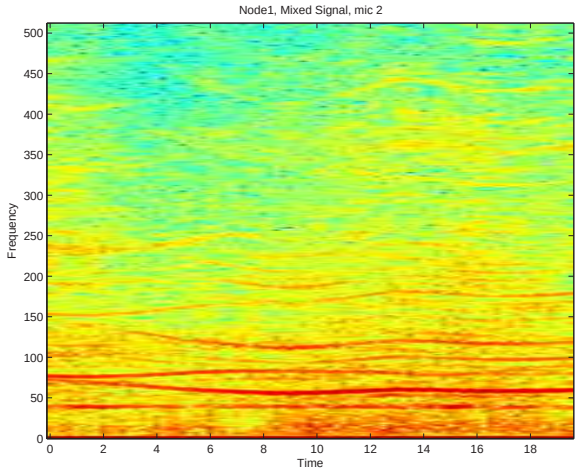


Figure 5. Spectrogram of recorded (mixed) signal at the center microphone.

5.1.2. Real Cases

The performance of the source separation algorithms are also evaluated on the real data consisting of multiple targets. We choose a data segment from one of the runs in the acoustic signature database that contains three moving targets. Figure 5 shows the spectrogram of the signal recorded at the center microphone of node 1 or first array. Figures 6(a) and (b) illustrate the spectrograms of the separated sources using the frequency domain ICA-based and nonstationary BSS methods, respectively. Visual evaluation of these results again indicates that the source separation capability of the nonstationary-based BSS method is superior to that of the frequency domain ICA-based method. More specifically, we notice that the separated sources using the latter method capture the harmonic structure of the sources and further they don't share common signatures as much as those in Figure 6(a) for the frequency domain ICA method. Consequently, the nonstationary BSS algorithm is used in the subsequent studies.

5.2. Classification Results

Four-class vehicle classification problem is considered here and one target from each of the four classes of vehicles is selected. The four classes are light tracked (LT), heavy tracked (HT), light wheeled (LW), heavy wheeled (HW). Classifier is a simple three-layer BPNN with thirty inputs, 45 neurons in the first hidden layer, 20 neurons in the second hidden layer, and four output neurons for four different vehicle classes. Features are extracted from the clean acoustic signature data and single-target runs. The data set can be divided into two subsets, training and validation. The training set consisted of 400 samples (100 samples from each class) and the validation set consisted of 160 samples. The classifier was trained using the samples in the training set for twenty different

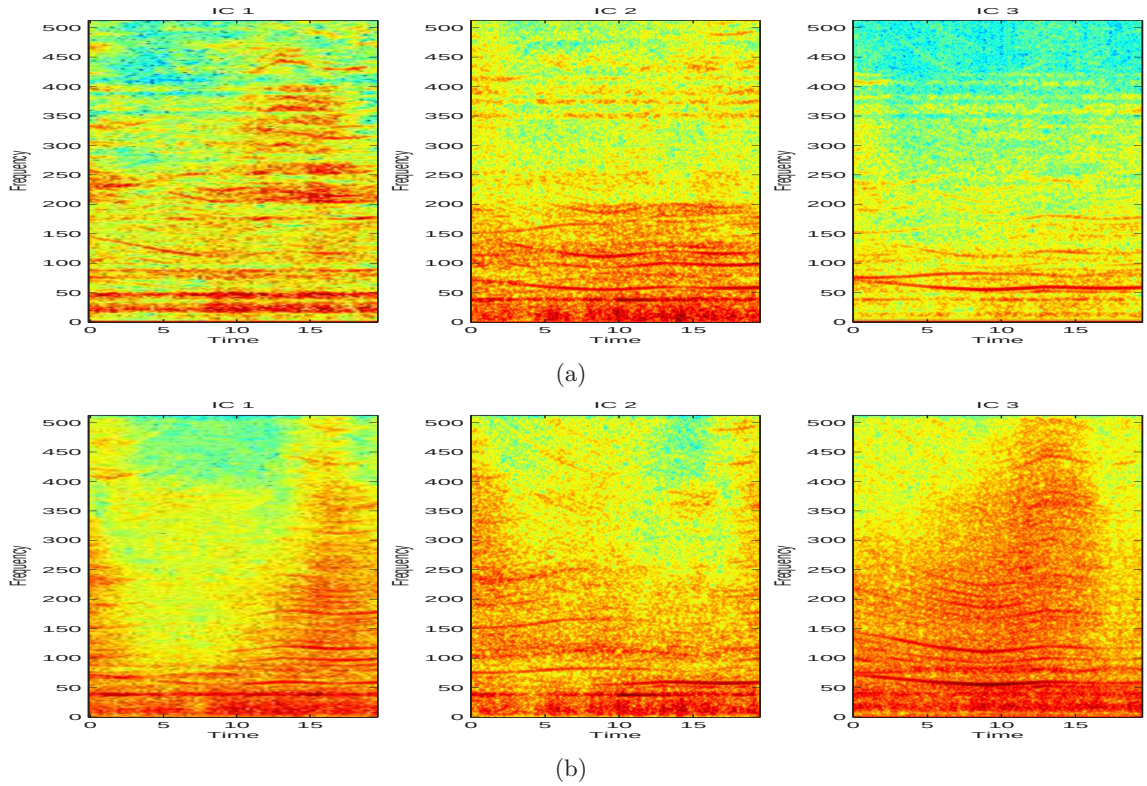


Figure 6. Separated sources using (a) frequency domain ICA-based method, and (b) Non-stationary BSS method.

random weight initializations and the network that gave minimum classification error on the validation set was chosen for subsequent testing. The correct classification rates on the training and the validation sets are 99.25% and 95%, respectively. The high correct classification rates demonstrate the discriminatory ability of the features for target classification.

Classification results on a challenging run that contains recorded signatures at three nodes (or arrays) are presented in this paper. This particular run contains one LW, three HTs and two HWs in the order of closest point of approach (CPA) to the three arrays. Classification is performed for every 10 second observation period based on the features extracted from this segment of the data. In the data segments that multiple classes of targets are detected, BSS is performed and then the separated sources are classified. For the data of this run, BSS had to be performed in time segments 300-400secs for node 1, 260-360secs for nodes 2 and 3. Tables 3-5 show the confusion matrices of the classifier for the data recorded by nodes (arrays) 1-3, respectively. From these tables, one can observe that, the correct classification rates for LW, HT and HW sources are 83.3% (10 out of 12), 84.4% (63 out of 75) and 91.6% (33 out of 36), respectively. The LT type vehicle was not present in this run and none of the other sources was misclassified as LT. For HT sources, 12 samples were misclassified, out of which 10 were misclassified as HW. This may be expected as HT and HW vehicles indeed share some common spectral and temporal features. Nevertheless, we were still able to achieve high correct classification rates on these two vehicles.

6. CONCLUSIONS

Two different source separation algorithms were studied for the separation of multiple target signatures. These methods rely on the frequency domain extension of the ICA algorithm and a joint diagonalization of the non-stationary spectra. A new subband-based spectral and temporal feature extraction method was also developed that extracts a small set of features that represent the pertinent spectral and temporal behavior of the sources

Table 3. Confusion matrix for Node 1, Run 508

	LW	HT	HW	LT
LW	2	0	0	0
HT	0	20	5	0
HW	0	0	12	0
LT	0	0	0	0

Table 4. Confusion matrix for Node 2, Run 508

	LW	HT	HW	LT
LW	4	1	0	0
HT	0	23	3	0
HW	1	0	10	0
LT	0	0	0	0

Table 5. Confusion matrix for Node 3, Run 508

	LW	HT	HW	LT
LW	4	1	0	0
HT	2	20	2	0
HW	2	0	11	0
LT	0	0	0	0

in an observation period. The results on both synthesized and real data cases revealed the usefulness of the nonstationary BSS method. Source separation using this BSS method was then performed on a real multiple target run that contains six sources of three different classes. Classification results using a BPNN classifier indicated promising results even on this challenging multiple target run. Although, BSS is able to separate vehicles of different classes, multiple vehicles of the same type cannot be discriminated using this method. This limitation had to be overcome by other ways of vehicle counting algorithms.

ACKNOWLEDGMENTS

This work is funded by Army SBIR-Phase II contract # DAAE30-03-C-1055. The data and technical support have been provided by the US Army TACOM-ARDEC, Picatinny Arsenal, NJ. The authors would like to thank Bob Wade and Myron Hohil for their invaluable suggestions and technical support.

REFERENCES

1. N. Srour, "Unattended Ground Sensors- A Prospective for Operational Needs and Requirements," *ARL Report Prepared for NATO*, October 1999.
2. T. Pham and M. Fong, "Real-time implementation of MUSIC for wideband acoustic detection and tracking," *Proc. of SPIE AeroSense'97: Automatic Target Recognition VII*, Orlando, FL, April 1997.
3. T. Pham and B. M. Sadler, "Wideband Array Processing Algorithms for Acoustic Tracking of Ground Vehicles," *ARL Technical Report*, Adelphi, MD, 1997.
4. F. Asano, S. Ikeda, M. Ogawa, H. Asoh, and N. Kitawaki, "Combined Approach of Array Processing and Independent Component Analysis for Blind Separation of Acoustic Signals", *IEEE Trans. on Speech and Audio Processing,* vol. 11, pp. 204-215, May 2003.
5. D.T. Pham and J.F. Cardosa, "Blind Separatin of Instantaneous Mixtures of Nonstationary Sources,", *IEEE Trans. Signal Processing*, vol. 49, pp. 1837-1848, Sept 2001.
6. D.T. Pham, C. Serviere, H. Boumaraf, "Blind Separation of Convolutive Audio mixtures Using Nonstationarity," Proc. of the *Intl. Symp. on Independent Component Analysis and Blind Signal Separation (ICA2003)*, April 2003, Nara, Japan.
7. D.T. Pham, "Joint Approximate Diagonalization of Positive Definite Hermitian Matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 22, pp. pp. 1136-1152.
8. S. Amari,"Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251-276, 1998.
9. A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol.7, pp.1129-1159, 1995.
10. A. Hyvarinen, J. Karhunen, and E. Oja, *Independent Component Analysis,* Wiley, 2001.
11. N. Murata, S. Ikeda, and A. Ziehe, "An Approach to Blind Source Separation Based on Temporal Structure of Speech Signals ", *Neurocomputing*, vol. 41, pp. 1–24, 2001.
12. S. Mallat, *A wavelet tour of signal processing*, Academic Press, 1998.