ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Signal Processing in Acoustics
Session 3aSP: Methods and Applications of Time-Frequency Analysis

3aSP5. Single snapshot spatial array processing using time-frequency distributions

Karim G. Sabra*

*Corresponding author's address: Georgia Tech, Atlanta, GA 30332, karim.sabra@me.gatech.edu

Several signal processing applications, such as spatial beamforming, rely on data collected by an array of sensor to estimate to enhance a weak signal in the presence of noise (e.g. ambient noise or clutter). The commonly applied eigenstructure subspace methods to this signal denoising problem assume stationary signals and require multiple snapshots to correctly estimate the covariance matrix of the array data. However, these multiple snapshots and stationarity requirements can be hard to meet in practical scenarios involving among others a rapidly moving source (which causes differential Doppler effect among sensors) or a single snapshot of the aspect-dependent scattering of an unknown target as measured by a monostatic sonar system (such as side-scan sonar). To handle these scenarios, we propose to form a generalized space-time-frequency covariance matrix from the single-snapshot data by computing Cohen’s class time-frequency distributions between all sensor data pairs. The eigenstructure of this space-time-frequency covariance matrix allows to enhance the localization of the signal of interest while spreading the noise power in the time-frequency domain. Hence, this approach is especially suited to handle nonstationary echoes of underwater target resonances that are highly localized in the time-frequency domain as demonstrated using numerical and at-sea data.

Published by the Acoustical Society of America through the American Institute of Physics
Introduction

The development of robust methodologies allowing for the concurrent detection, classification, and localization of underwater targets is a challenging problem with high operational importance for mine countermeasure activities (MCM). To this end, low-frequency (f~kHz) SONAR systems have been developed to enhance the detection of buried targets as well as the recognition of mine-like elastic targets by exciting their structural responses (or resonance signatures) [Lucifredi and Schmidt, 2006]. In this “structural acoustic regime” of frequencies, the scattered acoustic field from elastic targets includes both specular echoes (generated by geometric reflections from the target’s contour) as well as structural echoes, such as guided waves circumnavigating a thin shell [Zhang et al., 1992]. This elastic response has been suggested as a basis for target classification through acoustic “finger printing” using time-frequency analysis [Flandrin et al 1986; Cohen, 1995] or acoustic color representation [Williams et al., 2010].

But, due to radiation damping effects, especially when the target is (partially) buried, the amplitude of these structural echoes can be weak and thus hard to detect in the presence of surrounding clutter or ambient noise. When these structural echoes are recorded across a sufficiently large spatial aperture, it has been suggested that standard free-space back-propagation algorithms, i.e. based on point scatters’ radiation only (as used in Synthetics Aperture Sonar- or SAS- processing) could be used to numerically refocus these weak echoes onto their spatial origin to potentially better visualize and detect them [Lepage and Schmidt 2002, Lucifredi and Schmidt 2006, Williams et al., 2010]. However, due to their specific generation mechanisms, structural echoes of an elastic target can have a complex radiation pattern, causing the corresponding backscatter wavefronts to have different time-delay laws and frequency content than specular echoes have [Zhang et al., 1992, Anderson and Sabra, 2012]. Consequently structural echoes cannot readily be modeled as emanating from a collection of point scatterers radiating in free space contrary as commonly done for specular echoes. Thus, when compared to the specular echoes, these structural echoes typically appear defocused away from the actual target’s location on acoustic images generated using free-space back-propagation algorithms. This defocusing is especially detrimental for target recognition purposes when one attempts to image the spatial origin of weak structural echoes (compared to usually more energetic specular echoes) in the presence of ambient noise, reverberation, and clutter. Overall, this indicates a need for developing robust array signal processing methodologies to enhance the detection of non-stationary structural echoes of elastic targets.

Several signal processing applications, such as spatial beamforming, have been developed to enhance a weak signal in the presence of noise (e.g. ambient noise or clutter) by leveraging its coherence across the aperture of a receiver array. Eigenstructure subspace methods [Van Trees Ref, 2002]. But the commonly applied eigenstructure subspace methods to this signal denoising problem assume stationary signals and require multiple snapshots to correctly estimate the covariance matrix of the array data [Jensen et al., 2011]. However in several practical scenario involving active or passive SONAR measurements these requirements for multiple snapshots and stationarity are not met, thus preventing the direct implementation of commonly applied eigenstructure subspace methods For instance for typical operating conditions of SONAR systems mounted on moving platform (such as side-scan SONAR) (1) Doppler effects due to the relative motion of the source/receivers/target configuration can cause differential Doppler effect among sensors thus leading to non-stationarity of the received SONAR data, and (2) the scattered field of most elastic targets (e.g. cylindrical shells) is aspect-dependent; consequently, the received signals along the moving aperture vary for each SONAR’s illumination such that only a single-snapshot is available for a given source/receivers/target configuration.

To handle these scenarios, we propose to form a generalized space-time-frequency covariance matrix from single-snapshot data by computing Cohen’s class time-frequency distributions between all sensor data pairs [Cohen, 1995]. The eigenstructure of this space-time-frequency covariance matrix allows enhancing the localization of the signal of interest while spreading the noise power in the time-frequency domain. Hence, this approach is especially suited to handle nonstationary echoes of underwater target resonances that are highly localized in the time-frequency domain [Flandrin et al 1986; Zhang et al., 1992, Anderson and Sabra, 2012]. This approach is demonstrated experimentally to denoise the scattered field associated with the mid-frequency enhancement of a thin spherical shell measured along a synthetic SONAR aperture.
Theory

Let’s consider next the case of receiver array composed of N discrete elements. The signal recorded by each element \(i=1..N\) is denoted \(x_i(t)\). For instance, for the SONAR problem mentioned in the Introduction section, these signals would correspond to a single-snapshot of the target’s scattered field recorded on a N-elements receiver aperture for a given source broadcast. Furthermore, the following signal model is used for the SONAR signals: no assumption on the stationarity of the signals \(x_i(t)\) (\(i=1..N\)). Let’s assume that each broadband signals \(x_i(t)\) (\(i=1..N\)) each contain one main broadband signal component \(s_i(t)\) localized in the time-frequency plane (such as a structural echo of an elastic target) in the presence of additive noise \(n_i(t)\) such that \(x_i(t) = s_i(t) + n_i(t)\) (\(i=1..N\)). The exact time-frequency signature of this signal component \(s_i(t)\) may vary in the time-frequency plane due to their specific generation mechanism; but it is assumed that these broadband signal components \(s_i(t)\) have some degree of coherence in the time-frequency plane. For instance the signal components \(s_i(t)\) (\(i=2...N\)) could be a companded version of the signal component \(s_1(t)\) received on the first element, i.e. \(s_i(t) = s_1(\beta_i(t-t_i))\), where \(t_i\) and \(\beta_i\) correspond to the relative time and frequency shift of each signal \(s_i(t)\) with respect to \(s_1(t)\) (i.e. \(\beta_1=1, t_1=0\)) [Weiss, 1994]. A basis pulse for the signal \(s_1(t)\) could be a sinusoidal function modulated by a peaked Gaussian function (also known as Gabor wavelet or “time-frequency” atom) whose energy distribution is concentrated around a specific pixel of the time-frequency plane [Cohen, 1995]. This model for the signals \(x_i(t)\) (\(i=1..N\)) can be representative of the time-frequency evolution of a structural echo of an elastic target recorded across a bistatic aperture [Anderson and Sabra, 2012]. On the other hand the noise components \(n_i(t)\) [\(i=1..N\)] are assumed to be largely uncorrelated and have larger time—bandwidth products such that their effective noise power is distributed across the whole time-frequency plane. The objective here is to leverage the time-frequency coherence of the signal components \(s_i(t)\) (\(i=1..N\)) across the receiver aperture to enhance the detection of their localized time-frequency signatures by effectively reducing the contribution of the noise components \(n_i(t)\) (\(i=1..N\)).

The Cross Smooth Pseudo Wigner-Ville distribution (CSPWV) between two (complex) signals \(x_1(t)\) and \(x_2(t)\) is defined as [Cohen, 1995]

\[
WV_{x_1x_2}(t, f) = \int_{-\infty}^{+\infty} h(\tau) \int_{-\infty}^{+\infty} g(u-t) x_1(u+\tau/2) x_2^*(u-\tau/2) e^{-j2\pi f \tau} \, du \, d\tau
\]  

(1)

where \(t\) is the time variable and \(f\) the frequency. The function \(x_1(u+\tau/2) x_2^*(u+\tau/2)\) in the inner integrand represents the local cross-correlation of the analyzed signals \(x_1\) and \(x_2\) for a given symmetric time-shift \(\tau/2\). The separable kernel functions for time and frequency smoothing windows \(h(t)\) and \(g(t)\) (selected as symmetric Hann windows hereafter) are used to minimized crossed-terms interferences in the time-frequency plane [Cohen, 1995]. These smoothing functions allow to independently select the resolution of the CSPWV along the time and frequency axis. The CSPWV provides a mean to compute the coherence of two non-stationary signals along both the time and frequency axis [Cohen, 1995]. In practice a discrete version of the CSPWV is computed numerically. In this case, assuming that both discretized versions of the signals \(x_1\) and \(x_2\) contain each \(N_t\) time samples, the computed discretized CSPWV is a \(N_t \times N_t\) matrix where \(N_t\) is the number of frequency bins used to compute the discrete Fourier transform version for the discretized version of Eq. (1) [Cohen, 1995]. Eq. (1) can be used here to compute the discretized CSPWV between all pair-wise combinations of the received signals \(x_i(t)\), (\(i=1..N\)). The spatial time-frequency distribution (STFD) matrix of the \(N\) received signals \(x_i(t)\) is then constructed by concatenating all pair-wise discretized \(N_t \times N_t\) CSPWV matrices in the following fashion:

\[
STFD = \begin{bmatrix}
WV_{x_1x_1}(t, f) & WV_{x_1x_2}(t, f) & \cdots & WV_{x_1x_N}(t, f) \\
WV_{x_2x_1}(t, f) & WV_{x_2x_2}(t, f) & \cdots & WV_{x_2x_N}(t, f) \\
\vdots & \vdots & \ddots & \vdots \\
WV_{x_Nx_1}(t, f) & WV_{x_Nx_2}(t, f) & \cdots & WV_{x_Nx_N}(t, f)
\end{bmatrix}
\]  

(2)
This rectangular STFD matrix has the following dimensions \((NN_t)x(NN_f)\) (note that \(N_t \neq N_f\)). The block diagonal \(N_t \times N_f\) elements obtained when \(i=j\), \(\text{WV}_{x_i x_j}(t,f)\) \((i=1..N)\) correspond here to the discrete Smooth Pseudo Wigner-Ville distribution of each discretized signal \(x_i\). Furthermore this STFD matrix is hermitian since \(\text{WV}_{x_i x_j}(t,f) = \left(\text{WV}_{x_j x_i}(t,f)\right)^H\) based on the definition of the CSPWV in Eq. (1). This \((NN_t)x(NN_f)\) STFD matrix can be thought a generalization of the conventional \(N_xN\) covariance matrix array computed for a single frequency [Van Trees Ref, 2002]. However, an important distinction is that the rank of this STFD matrix is typically larger than one, although it was constructed from a single-snapshot of the received data \(x_i(t)\) \((i=1..N)\). Note that the conventional \(N_xN\) covariance array matrix would always be of rank one if only a single-snapshot is used [Van Trees Ref, 2002].

In previous work by Zhang et al. [2001], the STFD matrix was defined instead by concatenating the values of the pair-wise discretized CSPWV matrices for only one value of the time and frequency variables \(t\) and \(f\) (i.e. a single t-f pixel of the whole time-frequency plane) such that their STFD matrix had only of dimensions \(N_xN\). Zhang et al. [2001], then proposed to apply the same eigenstructure subspace methods to this square \(N_xN\) STFD matrix that were originally developed for the conventional \(N_xN\) covariance matrix of the array data (e.g. see [Jensen et al., 2011]). Various suggestions were made by Zhang et al. [2001], on how to optimally combined the information contained in the square \(N_xN\) STFD matrices for different t-f pixels to better extract signal components of interest contained in the received signals \(x_i(t)\), \((i=1..N)\) in the presence of additive noise. However the approach proposed by Zhang et al. [2001], does not intrinsically leverage the full time-frequency coherence of the received signals \(x_i(t)\), \((i=1..N)\) across all \(N_xN\)-time-frequency pixels.

We propose here to directly apply a singular-value decomposition to the large rectangular \(N_xN\) matrix, as defined in Eq. (2), to enhance the extraction of the signal components \(s_i(t)\) over the noise components \(n_i(t)\). The concept of decomposing a time-frequency distribution in this manner for signal denoising purposes was first done by Marinovich and Eichman [1985]. But their approach was only implemented to denoise the Wigner-Ville distribution of a single signal at a time; that is for a single \(N_xN\) matrix, \(\text{WV}_{x_i x_j}(t,f)\) \((i=1..N)\) of the block diagonal of the whole STFD matrix defined in Eq. (2). Instead here the singular-value decomposition SVD of the STFD matrix is given by (assuming \(N_x>N_f\) hereafter):

\[
\text{STFD} = UDV^H = \sum_{j=1}^{NN_f} \sigma_j u_j v_j^H. \tag{3}
\]

where \(D=\text{diag}(\sigma_j)\), \((j=1..NN_f)\) is a diagonal matrix and the singular values \(\sigma_j\) are ordered by decreasing magnitude. \(U\) and \(V\) are orthogonal matrices. Columns of \(U\) (or \(V\)) are the left (or right) singular vectors \(u_j\) (or \(v_j\)) of the STFD matrix. The expansion term \(\sigma_j u_j v_j^H\) of singular-value decomposition is named the \(j^{th}\) principal component of the STFD matrix. If the received signals \(x_i(t)\) \((i=1..N)\) follow the previously described signal model at the beginning of this section, the power of the noise components are distributed over a larger number of principal components than the power of the signal components. More specifically, if \(T(\delta)\) and \(B(\delta)\) are the average values of the effective duration and bandwidth of the signal components \(s_i(t)\) \((i=1..N)\) so that only a \(\delta\) fraction of the signal energy lay outside \(T(\delta)\) and \(B(\delta)\), the essential dimensionality of these signals in the time-frequency plane can be approximated by \(2T(\delta)B(\delta)\). It was experimentally observed, as first pointed out by Marinovich and Eichman [1985] for monocomponent signal, that in the absence of noise, the number of significant principal components \(J(\delta)\) that contain \((1-2\delta)\) fraction of the signal energy is approximately equal to the dimensionality of the signal \(J=2T(\delta)B(\delta)\). This estimation of \(J(\delta)\) was found to still hold for moderate noise level. Consequently a denoised approximation of the STFD can be obtained by truncating the singular value expansion in Eq. (3) to the first \(J(\delta)\) principal components.

Experimental Results

The denoising methodology introduced in the previous section is applied here to scattered field of a thin steel spherical shell measured by a monostatic SONAR configuration (see Fig. 1a). Experiments were conducted by the Naval Surface Warfare Center (Panama City, Florida) at their instrumented pond facility, which is a 14-m deep, 110-m long, and 80-m wide test-pool with a 1.5 m layer of sand on the bottom. The sound speed in the water was 1486 m/s. The experimental collection methods and apparatus is discussed in details in a previous study [Williams et al., 2010] conducted by the team of researchers who shared these acoustic backscatter data with the author of this
article. In short, the source and receiver array were mounted in quasi-monostatic configuration on a rigid tower frame and were located at a depth of 10 m (see Fig. 1b). The transmitter and receiver array were mounted on a panel-tilted at 20° angle towards the bottom and were separated horizontally by about half a meter. The receiver array, composed of six hydrophones, had a 10 cm horizontal aperture and 1 m vertical aperture. The received signals by the 6 elements were added coherently (broadside summation) to minimize scattering interference from the water/air boundary. The source-receiver tower was moved in 2.5 cm increments along a 20 m long rail to create a synthetic aperture and collect backscattering measurements of the elastic targets (see Fig. 2). The transmitter maintained a horizontal beamwidth greater than 40° over the entire frequency band to allow for SAS processing. At the point of closest approach—corresponding to a cross-range of 0.8 m here—the horizontal distance between the source-receiver tower and targets (both centered on approximately the same location) was close to 10 m. From these measurements, the grazing angle for the ray drawn from the centers of the transmitter and receiver array to the center of each target was approximately 21° (see Fig. 1b).

The first target studied was a hollow 59.9 cm diameter steel spherical shell (density = 5773 kg/m³, compressional velocity = 7970 m/s and shear velocity = 3020 m/s). The sphere was excited by a linear frequency modulated (LFM) waveform ranging from 12 kHz to 28 kHz. Fig. 2 displays the backscatter signals measured by the source-receiver tower while moving in 2.5 cm increment along the cross-range axis. The first (or second) main group of wavefronts centered around 15.5 ms (or 16.5 ms) at the closes point of approach (i.e. cross-range of 10) and 16.8 ms, correspond to the specular echo (or first structural echo) of the spherical shell. The first structural echo is dominated by the constructive interference of antisymmetric guided waves circumnavigating the shell’s perimeter along reciprocal paths in clockwise and counterclockwise direction (also referred to as mid-frequency enhancement effect) [Zhang, 1992; Anderson and Sabra 2012]. In this waveguide geometry, both specular and structural echoes are convolved with the multipath arrival structure of the shallow water waveguide. Based on a simple ray approximation (see Fig. 1b), these multipath arrivals shown on Fig. 2 can be interpreted as acoustic energy traveling along four different paths between the source and receiver arrays, namely (1) the first direct reflection path (Source-Shell-Receiver path), followed by (2) two reciprocal paths of equal length (thus recorded simultaneously by the receivers) interacting only once with the bottom (Source-Shell-Bottom-Receiver and Source-Bottom-Shell-Receiver paths) and finally (3) a path interacting twice with the bottom (Source-Bottom-Shell-Bottom-Receiver) thus having a weaker amplitude than the previous paths. Note, that the angles of incidence on the shell of these bottom-interacting multipaths effectively introduce bistatic scattering paths even though the source and receiver are arranged in a quasi-monostatic geometry [Williams et al. 2010, Anderson and Sabra, 2012].

Due to the azimuthal symmetry of the sphere, the free-space response of a spherical shell is independent of the source/receiver position for a monostatic configuration. Hence the specular and structural echoes of the shell associated the first direct path (which can be well approximated by the shell’s free-space response Williams et al. 2010]) is expected to be the same for each cross-range position along the rail (see Fig. 1a), up to a propagation delay due to the varying distance between the shell and the source/receiver (see Fig. 2). However, due to the finite horizontal beamwidth of the transmitter, the shell’s ensonification was poor towards the ends of the receiver’s track which resulted in a very low signal-to-noise ratio (SNR) of the backscattered data for cross-ranges above 16 m and below 5.5 m here (see Fig. 2). For instance on Fig. 3a and Fig. 3c, the specular (around 16.2 ms) and first structural (around 17 ms) echoes recorded at a cross-range of 13.5 m are clearly visible on the received signal and its time-
frequency representation obtained using the smooth pseudo Wigner-Ville transform as defined in Eq. (1) (by setting $x_1 = x_2$). On the other hand, Fig. 3b and Fig. 3d the specular (around 16.8ms) and first structural (around 17.8ms) echoes recorded at a cross-range of 15m can barely be identified from the received signal and its time-frequency representation, and a strong clutter return appears around 18.4ms.

**FIGURE 2.** Evolution of the envelope (in logarithmic scale) of the spherical shell's backscatter vs. receiver's cross-range. The amplitudes were normalized with respect to the maximum displayed value.

**FIGURE 3.** Backscatter time-series of the spherical shell recorded at cross-ranges (a) 13.5m and (b) 15m. (c-d) The time-frequency distribution, plotted below its corresponding time-series below each signal, was computed using a Smooth Pseudo Wigner-Ville transform (see Eq. (1)).

The methodology presented in the Theory section was applied here to attempt to enhance the effective SNR of the backscattered data recorded towards the ends of the receiver’s track where target’s ensonification was poor, i.e. for cross-ranges above 16m and below 5.5m here (see Fig. 2). Enhancing the SNR of these backscattered data at the track’s extremities, would be advantageous to improve the detection of the structural echo as well as extending the effective available cross-range aperture to potential enhance the focusing resolution of SAS processing of these backscattered data. To do so, a section of the wavefront associated with the first structural echo (corresponding to the second main wavefront on Fig. 2) and having weak SNR was first extracted from the corresponding backscattered data by appropriately time-gating $N=41$ received signals recorded along a 1m long cross-range aperture centered at a cross-range of 15m (see Fig. 4.a). The STFD matrix of these $N=41$ truncated signals was then computed using Eq. (1-2). Subsequently, a denoised version of the STDF matrix was estimated using its first principal component (i.e. $\sigma_{\mu_1}$ using notations from Eq. (3)) as most of the power of wavefront  corresponding to the structural echo was found to be concentrated in this first principal component. By analogy to Eq. 2, each of the $N$ block diagonal matrix of size $N_t \times N_f$, noted $W(1)_{x_1}(t, f)$, of the first principal component of the STFD contain a denoised version of the discrete Smooth Pseudo Wigner-Ville distribution of each the original truncated signal noted $x_i(t)$, $(i=1..41)$. Hence, computing the first moment about the frequency variable $f$ of these denoised time-frequency distributions $W(1)_{x_1}(t, f)$ yielded a denoised version of the envelope of the structural echo.
wavefront (as shown on Fig 4.b) was Cohen 1995]. Overall Fig. 4b shows an average SNR improvement of more
than 15dB for the enveloped of the denoised wavefront associated with the direct arrival of the structural echo (and a
weaker multipath arrival) when compared to original section of the noisy wavefront displayed in Fig. 4a.

Conclusion

A subspace-based denoising method based on principal component analysis of space-time-frequency distributions
was introduced for single-snapshot array data. This method was applied to the case of N received broadband
containing each one main broadband (nonstationary across receivers) signal component (e.g. elastic response of
underwater targets) localized in the time-frequency plane. This approach allows to automatically leverage the time-
frequency coherence of the signal components across the receiver aperture. The applicability of this method for
denoising backscattered data of a thin spherical shell was demonstrated experimentally, thus enhancing the detection
of weak structural echoes for MCM purposes.

FIGURE 4. (a) Envelope (in logarithmic scale) of a noisy section of the wavefront associated with the first structural echo
(corresponding to the second main wavefront on Fig. 2) extracted from the original backscattered data (see Fig. 2) by
appropriately time-gating N=41 received signals recorded along a 1m long cross-range aperture centered at a cross-range of
15m. (b) Denoised version of the envelope of the N=41 waveforms shown in (a) obtained from principal component analysis of
their space-time-frequency distribution (STFD).

ACKNOWLEDGMENTS

This work was supported by the Office of Naval Research (Code 321) under Contract No. N00014-08-1-0087.
The authors are grateful to Dr. Joseph Lopes and Dr. Kevin Williams for granting to the experimental data.

REFERENCES

America, vol. 120, no. 6, pp. 3566–3583.