

MSICA: MULTI-SCALE SIGNAL DECOMPOSITION BASED ON INDEPENDENT COMPONENT ANALYSIS WITH APPLICATION TO DENOISING AND RELIABLE MULTI-CHANNEL TRANSMISSION

Abolfazl Hajisami and Dario Pompili

Dept. of Electrical and Computer Engineering, Rutgers University–New Brunswick, NJ, USA

Abstract – Multi-scale decomposition is a signal description method in which the signal is decomposed into multiple scales, which has been shown to be a valuable method in information preservation. Much focus on multi-scale decomposition has been based on scale-space theory and wavelet transform. In this article, a new powerful method to perform multi-scale decomposition exploiting Independent Component Analysis (ICA), called MSICA, is proposed to translate an original signal into multiple statistically independent scales. It is proven that extracting the independent components of the even and odd samples of a digital signal results in the decomposition of the same into approximation and detail. It is also proven that the whitening procedure in ICA is equivalent to a filter bank structure. Performance results of MSICA in signal denoising are presented; also, the statistical independency of the approximation and detail is exploited to propose a novel signal-denoising strategy for multi-channel noisy transmissions aimed at improving communication reliability by exploiting channel diversity.

Keywords – Channel Diversity, Independent Component Analysis, Multi-scale Decomposition, Wavelet Transform.

1. INTRODUCTION

Overview: Multi-scale decomposition is an invaluable tool in digital signal processing with applications such as those in [1, 2, 3, 4, 5], where an original signal is decomposed into a set of signals, each of which provides information about the original signal at a different *scale*. A major signal-processing task where multi-scale decomposition has been shown to be very useful is denoising, based on the intuition that information pertaining to the noise would be accurately characterized in certain scales that are separate from the scales of the signal. The main literature works in multi-scale decomposition have focused on scale-space decomposition [6, 7, 8, 9, 10], empirical mode decomposition [11, 12, 13], and wavelet transform [14, 15, 16, 17, 18].

In scale-space theory [19], a signal is decomposed into a single-parameter family of n signals with a progressive decrease in fine scale signal information between successive scales. This allows analyzing signals at coarser scales without the influence of finer scales such as those pertaining to noise. Knowing this, one can employ scale-space theory to suppress noise by performing scale-space decomposition on the signal and then treating one of signals at a coarser scale as the noise-suppressed signal. However, selecting the scale that represents the noise-suppressed signal can be challenging. Moreover, noise suppression using scale-space theory does not facilitate the fine-grained noise suppression at the individual scales, which limits its overall flexibility in striking a balance between noise suppression and signal structural preservation [20].

In Discrete Wavelet Transform (DWT), the original signal is decomposed into *approximation* and *detail* by

passing the signal through a low-pass filter and high-pass filter, respectively, followed by a downsampling by a factor of 2. This results in a decomposition of the signal into different scales, which can be considered as low and high frequency bands. Multi-scale decomposition by wavelet transforms has a number of advantages over the scale-space decomposition and empirical mode decomposition [20]. First, since the signal information at one scale is not contained in another scale, signal information at different scales are better separated in the wavelet domain. Second, scale selection when performing noise suppression using wavelet decomposition is less critical than that for noise suppression using scale-space decomposition as all the scales are considered in noise suppression using wavelet decomposition as opposed to a single scale as done in scale-space decomposition. However, there are a number of limitations pertaining to noise suppression using wavelet transform [20]. For instance, signals processed using wavelet transforms can exhibit oscillation artifacts related to wavelet basis functions used in the wavelet transform, which is particularly noticeable in low Signal-to-Noise Ratio (SNR) regimes. Moreover, in DWT the approximation and detail are not statistically independent, which may cause a poor performance in signal denoising.

Motivation and Approach: Given these limitations of both space-scale and wavelet decomposition in terms of signal denoising, we were motivated to explore alternative approaches. We investigate the problem of decomposing a signal into multiple scales from a different point of view, i.e., we propose a new approach that takes a statistical perspective on multi-scale decomposition according to which a signal is considered as a mix-

ture of statistically independent signals, each characterizing signal information at a different scale. Having this perspective in multi-scale decomposition can be beneficial in signal denoising for two reasons. First, since most of the signal information in one scale is not included in the other scales, such decomposition provides the advantage of noise suppression at the individual scales in order to trade off noise suppression for signal-quality preservation. Second, since the noise signal is statistically independent from the original signal, by decomposing the noisy signal into statistically independent scales, the noise is expected to be separated in finer scales.

Our Contribution: Given this motivation and perspective, we propose a new method for Multi-Scale decomposition exploiting Independent Component Analysis (ICA), called MSICA, in which the original digital signal is decomposed into approximation and detail with statistically independent components. Specifically, we extract two correlated signals from the original signal and apply a linear transformation to the extracted signals so as to decompose the original signal into multiple scales. Since we need a suitable transform to decompose the original signal into statistically independent components, we consider our problem as a Blind Source Separation (BSS) problem in which the extracted signals are considered as the *observations* of the source separation problem. To relate this problem with the concept of multi-scale decomposition, we introduce an equivalent filter-bank structure for the proposed method, which is similar to the structure introduced in [14] for the DWT implementation. We also propose a method for multi-channel transmission in which MSICA outperforms common wavelet transforms in denoising of the received signal. We show that if the even and odd samples of the original signal are transmitted through two Additive White Gaussian Noise (AWGN) channels, MSICA is able to extract and filter out the noise from the noisier channel. This key property of MSICA—which exploits channel diversity and generalizes to the case in which more than two channels are available—can be used to increase the transmission efficiency in noisy communication channels, although the marginal return diminishes as the number of channels increases. It should be noted that, although single-channel ICA has been used in previous works [21, 22, 23, 24, 25] (including the spatial case of using even and odd samples), in this work single-channel ICA has been studied as a technique for signal decomposition into statistically independent approximation and detail and its performance in denoising has been compared with other wavelet transforms.

Article Organization: In Section 2, we provide some background on BSS and ICA. In Section 3, we propose our ICA-based transform for multi-scale decomposition. In Section 4, we examine the performance of MSICA in signal denoising and show how to increase transmission efficiency when multiple (noisy) channels are available. Finally, in Section 5, we draw the main conclusions and wrap up the article by discussing future work.

2. BLIND SOURCE SEPARATION

In BSS, a set of mixtures of different source signals is available and the goal is to separate the source signals when we have no information about the mixing system or the source signals (hence the name “blind”) [26, 27].

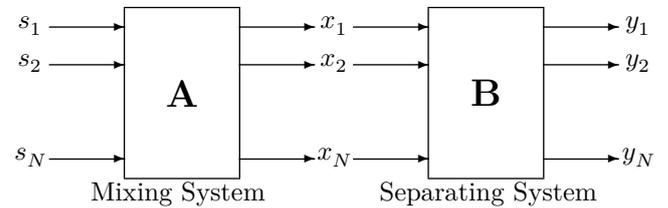


Fig. 1 – Mixing and separating systems in Blind Source Separation (BSS).

As in Fig. 1, the mixing and separating systems can be represented as,

$$\begin{aligned} \mathbf{x}(n) &= \mathbf{A}\mathbf{s}(n), \\ \mathbf{y}(n) &= \mathbf{B}\mathbf{x}(n), \end{aligned} \quad (1)$$

where $\mathbf{s}(n) = [s_1(n), \dots, s_N(n)]^T$ is the vector of sources that are mixed by the mixing matrix \mathbf{A} and create the observations vector $\mathbf{x}(n) = [x_1(n), \dots, x_N(n)]^T$. Let \mathbf{A} be a square matrix ($N \times N$) of full column rank, which means that the number of sources is equal to the number of observations and that the sources are linearly independent. The goal of BSS is to find the separating matrix \mathbf{B} such that $\mathbf{y}(n) = [y_1(n), \dots, y_N(n)]^T$ is an estimation of the sources.

A method to solve the BSS problem is via ICA, which exploits the assumption of source independence and estimates \mathbf{B} such that the outputs y_i 's be statistically independent [28]. As studied in [28, 29], the ICA can be performed by two steps: 1) whitening (or decorrelating) and 2) rotation. To illustrate the ICA model, we consider two independent components, s_i , $i = 1, 2$, with a uniform distribution,

$$p(s_i) = \begin{cases} 1 & \text{if } |s_i| \leq 0.5, \\ 0 & \text{otherwise,} \end{cases} \quad (2)$$

where the joint density of s_1 and s_2 is uniform on a square, as illustrated in Fig. 2(a). This follows from the definition that the joint density of two independent variables is the product of their marginal densities. Let us now mix these two independent components by the following mixing matrix \mathbf{A} ,

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}}_{\mathbf{A}} \begin{bmatrix} s_1 \\ s_2 \end{bmatrix}, \quad (3)$$

where the mixed variables x_1 and x_2 have a uniform distribution on a parallelogram, as shown in Fig. 2(b). Note that x_1 and x_2 are not independent anymore. To show this consider whether it is possible to predict the value of one of them, say x_2 , from the value of the other; it is clear that if x_1 attains one of its maximum or minimum values, then this completely determines the value

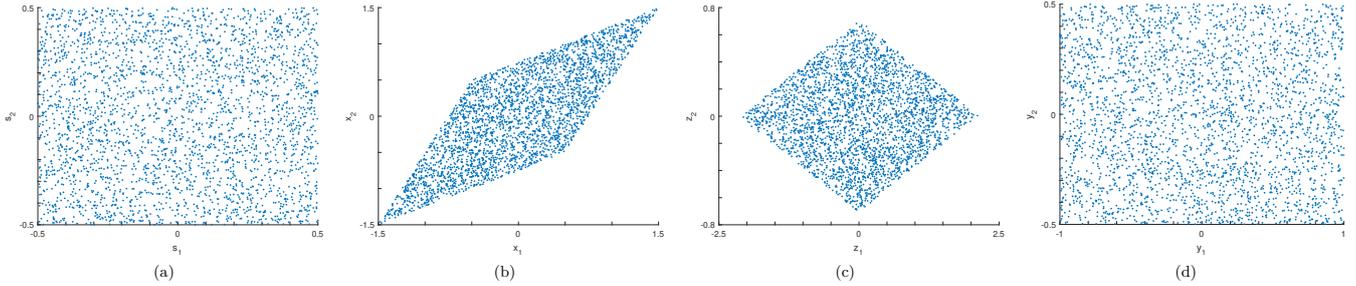


Fig. 2 – Illustration of Independent Component Analysis (ICA) algorithm. (a) Joint distribution of the independent components s_1 and s_2 with uniform distribution; (b) Joint distribution of the observed mixtures x_1 and x_2 ; (c) Joint distribution of the whitened mixtures, z_1 and z_2 ; (d) Joint distribution of the independent output components, y_1 and y_2 , as determined by the ICA.

of x_2 . However, the situation for variables s_1 and s_2 is different: from Fig. 2(a) it is clear that knowing the value of s_1 does not help predict the value of s_2 .

The problem of source separation is now to estimate the mixing matrix \mathbf{A} and multiply its inverse ($\mathbf{B} = \mathbf{A}^{-1}$) to the vector of the mixtures to retrieve the sources (s_1 and s_2). As studied in [28, 29], the ICA can be considered as a two-step procedure where, in the first step, the mixed data gets whitened (uncorrelated) by multiplying the whitening matrix by the vector of mixtures, i.e.,

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \mathbf{W} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \quad (4)$$

and then, in the second step, the independent components are separated by applying a rotation matrix, i.e.,

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \mathbf{R} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}. \quad (5)$$

Fig. 2(c) illustrates the effect of whitening in which the data in Fig. 2(b) has been whitened. Also, Fig. 2(d) shows how rotating the whitened data can return the statistical independency to the outputs and recover the independent components. Therefore, the separation matrix \mathbf{B} can be considered as the product of the whitening and rotation matrices, i.e.,

$$\mathbf{B} = \mathbf{R}\mathbf{W}, \quad (6)$$

where \mathbf{W} is the whitening matrix and \mathbf{R} is the rotation matrix. Note that in the case whitened components are statistically independent, the rotation matrix \mathbf{R} will be the identity matrix and no rotation will be needed.

3. MSICA: MULTI-SCALE DECOMPOSITION BY INDEPENDENT COMPONENT ANALYSIS

Generally, neighboring/consecutive samples of a common signal are highly correlated and differ slightly from each other. This slight difference of neighboring samples comes from the details of the signal. If we consider the detail scale of the original signal to be statistically independent from the approximation scale, it is expected

that decomposing the neighboring samples of the signal into their independent components would decompose the signal into its approximation and detail. Given this motivation, we propose an ICA-based method for multi-scale decomposition in which the approximation and details are statistically independent. Our algorithm consists of two steps: 1) extracting the observation signals from the original signal and 2) decomposing the original signal into approximation and detail by applying a linear transformation to the observation signals. Suppose that $x(n)$ is a Wide Sense Stationary (WSS) signal. We consider even and odd samples of $x(n)$ as $x_1(n)$ and $x_2(n)$, respectively, i.e.,

$$x_1(n) = x(2n), \quad x_2(n) = x(2n - 1). \quad (7)$$

We prove that, if $x_1(n)$ and $x_2(n)$ are the observations signals (mixtures) of the ICA, the outputs of the ICA will be the approximation and detail of $x(n)$, which are statistically independent. We define the observation vector \mathbf{x} as,

$$\mathbf{x} = \begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \begin{bmatrix} x(2n) \\ x(2n - 1) \end{bmatrix}, \quad (8)$$

where the even and odd samples of the original signal are the first and second rows of the observation vector, respectively. If we apply the ICA to the observation vector, the output is,

$$\mathbf{y} = \mathbf{B}\mathbf{x} = \begin{bmatrix} y_1(n) \\ y_2(n) \end{bmatrix}, \quad (9)$$

where \mathbf{B} is the separation matrix estimated by the ICA, and $y_1(n)$ and $y_2(n)$ are statistically independent.

To prove our claim, we need to show that $y_1(n)$ and $y_2(n)$ are the approximation and detail of the original signal. As explained in Section 2, ICA involves two steps: 1) whitening (or decorrelating) and 2) rotation, and the separation matrix \mathbf{B} can be considered as the product of the whitening and rotation matrices (i.e., $\mathbf{B} = \mathbf{R}\mathbf{W}$). Now, let us consider the whitened vector \mathbf{z} and the whitening matrix \mathbf{W} as follows,

$$\mathbf{z} = \begin{bmatrix} z_1(n) \\ z_2(n) \end{bmatrix} = \mathbf{W}\mathbf{x}, \quad \mathbf{W} = \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix}. \quad (10)$$

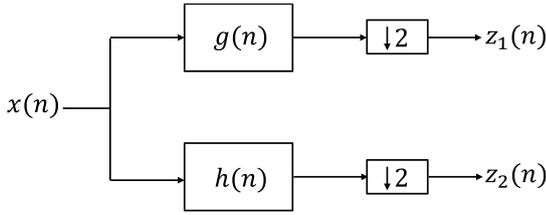


Fig. 3 – Whitening process in ICA as a filter-bank structure.

Using (8) and (10), we can write,

$$\begin{aligned} z_1(n) &= w_{11}x_1(n) + w_{12}x_2(n), \\ z_2(n) &= w_{21}x_1(n) + w_{22}x_2(n). \end{aligned} \quad (11)$$

Also, $x_1(n)$ and $x_2(n)$ can be written as,

$$\begin{aligned} x_1(n) &= x(2n) = [x(n) * \delta(n)] \downarrow 2, \\ x_2(n) &= x(2n-1) = [x(n) * \delta(n-1)] \downarrow 2, \end{aligned} \quad (12)$$

where $*$ and $\downarrow 2$ denote convolution and downsampling by a factor of 2, respectively, and $\delta(n)$ is the unit impulse signal. Using (11) and (12), we can rewrite $z_1(n)$ and $z_2(n)$ as,

$$\begin{aligned} z_1(n) &= w_{11}x(2n) + w_{12}x(2n-1) \\ &= w_{11}[x(n) * \delta(n)] \downarrow 2 + w_{12}[x(n) * \delta(n-1)] \downarrow 2 \\ &= [x(n) * (w_{11}\delta(n) + w_{12}\delta(n-1))] \downarrow 2, \\ z_2(n) &= w_{21}x(2n) + w_{22}x(2n-1) \\ &= w_{21}[x(n) * \delta(n)] \downarrow 2 + w_{22}[x(n) * \delta(n-1)] \downarrow 2 \\ &= [x(n) * (w_{21}\delta(n) + w_{22}\delta(n-1))] \downarrow 2. \end{aligned} \quad (13)$$

Note that, if we consider the low-pass filter $g(n)$ and high-pass filter $h(n)$ as,

$$\begin{aligned} g(n) &= w_{11}\delta(n) + w_{12}\delta(n-1), \\ h(n) &= w_{21}\delta(n) + w_{22}\delta(n-1), \end{aligned} \quad (14)$$

(13) can be rewritten in a simpler form as,

$$\begin{aligned} z_1(n) &= [x(n) * g(n)] \downarrow 2, \\ z_2(n) &= [x(n) * h(n)] \downarrow 2. \end{aligned} \quad (15)$$

From (15), we observe that the whitening process can be modeled as a filter-bank structure, as shown in Fig. 3. Now, we need to show that $g(n)$ and $h(n)$ are indeed low-pass and high-pass filters. In order to do so, we consider the covariance matrix of \mathbf{x} as follows,

$$\mathbf{C}_x = E\{\mathbf{xx}^T\} = \mathbf{Q}\mathbf{D}\mathbf{Q}^T, \quad (16)$$

where \mathbf{Q} is an orthogonal matrix of eigenvectors and \mathbf{D} is a diagonal matrix of eigenvalues. Interestingly, the covariance matrix of $\mathbf{z} = \mathbf{Q}^T\mathbf{x}$ can be written as,

$$\mathbf{C}_z = E\{\mathbf{zz}^T\} = E\{\mathbf{Q}^T\mathbf{xx}^T\mathbf{Q}\} = \mathbf{Q}^TE\{\mathbf{xx}^T\}\mathbf{Q}. \quad (17)$$

Given (16) and (17), we can write,

$$\mathbf{C}_z = \underbrace{\mathbf{Q}^T\mathbf{Q}}_{\mathbf{I}}\mathbf{D}\underbrace{\mathbf{Q}\mathbf{Q}^T}_{\mathbf{I}} = \mathbf{D}. \quad (18)$$

Since \mathbf{D} is a diagonal matrix, we conclude that multiplying \mathbf{Q}^T by the observation vector \mathbf{x} results in uncorrelated outputs. Hence, the whitening matrix \mathbf{W} can be considered to be equal to \mathbf{Q}^T .

Now, we need to find the elements of matrix \mathbf{W} so as to determine finally the $g(n)$ and $h(n)$ filters. Since $x(n)$ is a WSS signal, we have,

$$\begin{aligned} R_x(0) &= E\{x^2(2n)\} = E\{x^2(2n+1)\} = \sigma_x^2, \\ R_x(1) &= E\{x(2n)x(2n-1)\} = \sigma_x^2\rho. \end{aligned} \quad (19)$$

Hence, with regard to (8), we can recast \mathbf{C}_x as,

$$\mathbf{C}_x = E\{\mathbf{xx}^T\} = \begin{bmatrix} \sigma_x^2 & \sigma_x^2\rho \\ \sigma_x^2\rho & \sigma_x^2 \end{bmatrix} = \sigma_x^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}, \quad (20)$$

where the eigenvalues and eigenvectors of \mathbf{C}_x are,

$$\begin{aligned} \lambda_1 &= \sigma_x^2(1+\rho) & \lambda_2 &= \sigma_x^2(1-\rho), \\ \mathbf{q}_1 &= \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} & \mathbf{q}_2 &= \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}. \end{aligned} \quad (21)$$

Hence, the whitening matrix \mathbf{W} can be presented as,

$$\mathbf{W} = \mathbf{Q}^T = [\mathbf{q}_1 \quad \mathbf{q}_2]^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}. \quad (22)$$

Comparing (14) and (22), $g(n)$ and $h(n)$ can be written as,

$$\begin{aligned} g(n) &= \frac{1}{\sqrt{2}}\delta(n) + \frac{1}{\sqrt{2}}\delta(n-1), \\ h(n) &= -\frac{1}{\sqrt{2}}\delta(n) + \frac{1}{\sqrt{2}}\delta(n-1). \end{aligned} \quad (23)$$

From (23), it is clear that $g(n)$ and $h(n)$ are a low and high-pass filter, respectively, as we wanted to demonstrate. Hence, we can conclude that the whitening process in the ICA (presented as a filter-bank structure in Fig. 3) decomposes the observation signals into uncorrelated approximation and detail. Moreover, the rotation process in the ICA makes sure that the approximation and the detail are statistically independent¹. Hence, if the even and odd samples of a one-dimensional signal are considered as the observations of the ICA, we ensure the approximation and detail to be statistically independent. As we will see in the next section, the statistical independency between the approximation and detail can be very beneficial in signal denoising, especially when the even and odd samples are transmitted through different (noisy) channels.

Fig. 4 showcases a signal decomposition by different wavelet transform and MSICA, where a Piece-Regular signal is decomposed into the approximation and detail. As shown in Fig. 4(h), like all the other wavelet transforms, MSICA is also able to decompose the original signal into approximation and detail, where the approximation and detail coefficients contain the low and high-frequency components, respectively. As said earlier, to

¹Based on our simulations, the separation matrix was always close to (22), which means that the rotation matrix was always close to the identity matrix.

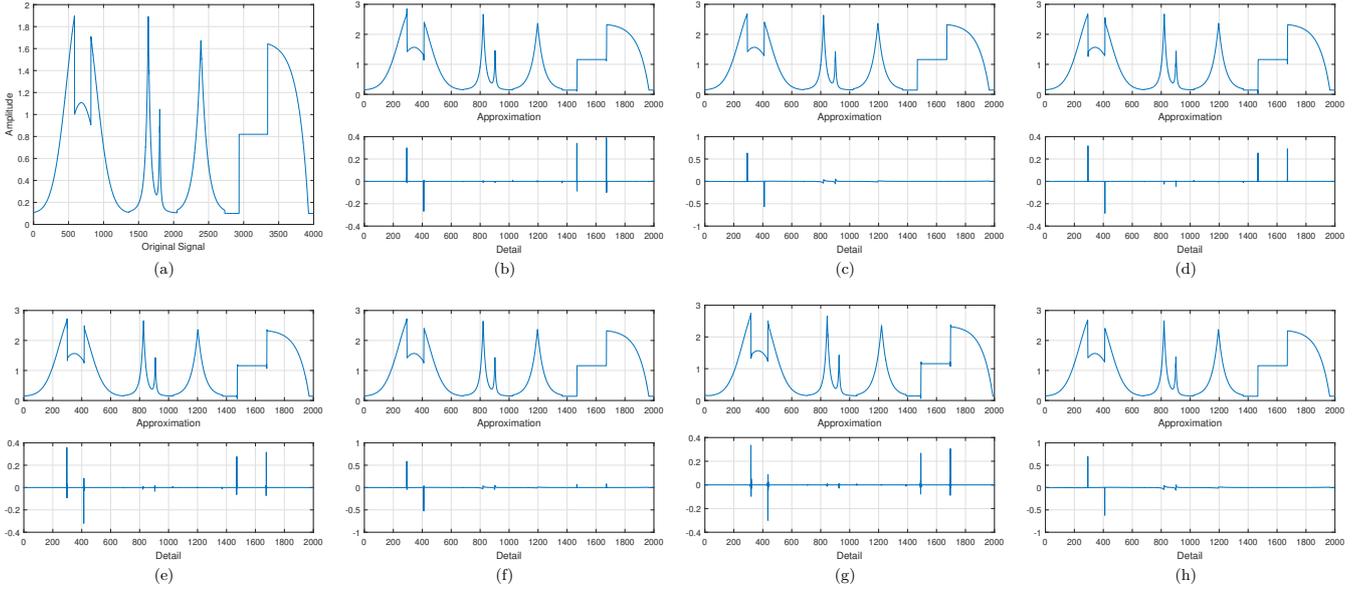


Fig. 4 – Comparison of MSICA with different wavelet transforms in decomposing a Piece-Regular signal. (a) Original Signal. Approximation and detail by (b) Daubechies 3 wavelet; (c) Haar wavelet; (d) Biorthogonal 2.2 wavelet; (e) Coiflets 4 wavelet; (f) Fejer-Korovkin 4 wavelet; (g) discrete Meyer wavelet; and (h) our proposed MSICA.

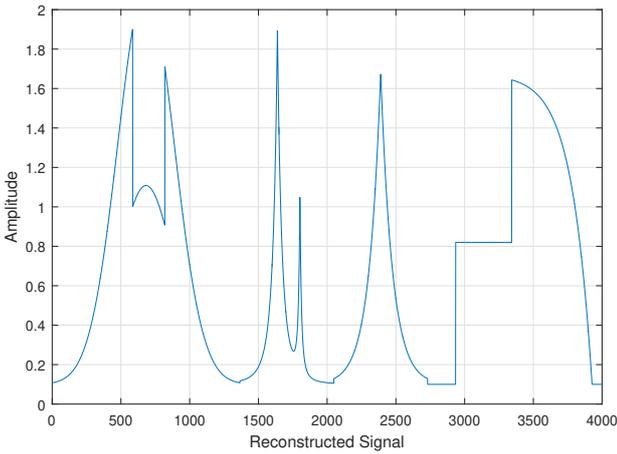


Fig. 5 – Reconstructed signal by MSICA using the approximation and detail in Fig. 4(h).

reconstruct the original signal we need to multiply the inverse of the separation matrix (\mathbf{B}^{-1}) by the vector of approximation and detail \mathbf{y} so as to obtain the observation vector \mathbf{x} and reconstruct the original signal $x(n)$ from \mathbf{x} , i.e.,

$$\begin{bmatrix} x_1(n) \\ x_2(n) \end{bmatrix} = \mathbf{x} = \mathbf{B}^{-1}\mathbf{y}, \quad (24)$$

$$x(n) = (x_1(n) \uparrow 2) + (x_2(n) \uparrow 2) * \delta(n+1), \quad (25)$$

where $x_i(n) \uparrow 2$, $i = 1, 2$, denotes the upsampling of $x_i(n)$ by a factor of 2. Fig. 5 shows the reconstructed signal using (24) and (25). This figure shows that MSICA can successfully reconstruct the original signal from the approximation and detail obtained in the decomposition procedure.

4. MSICA FOR SIGNAL DENOISING

We discuss now how MSICA can be beneficial in signal denoising. Specifically, we compare MSICA with the other wavelet transforms and show how MSICA can suppress the noise via a simple wavelet thresholding. We also show that, in the case of multi-channel transmission, MSICA outperforms the other wavelet transforms and is able to extract and filter out the noise of the noisier communication channel by exploiting channel diversity.

Let us assume that the original signal is passed through an AWGN channel, the noisy output signal is then,

$$r(n) = x(n) + w(n), \quad n = 1, \dots, N, \quad (26)$$

where $x(n)$ is the original signal and $w(n)$ is the AWGN with zero mean and variance of σ_w^2 . The goal of signal denoising is to remove the noise and obtain an estimate $\hat{x}(n)$ of $x(n)$ that minimizes the Mean Squared Error (MSE), defined as,

$$\text{MSE}(\hat{x}) = \frac{1}{N} \sum_{n=1}^N (\hat{x}(n) - x(n))^2. \quad (27)$$

Note that the model presented in (26) is not general since the noise may be non-additive, and the relation between the noisy observed signal and the original signal may be stochastic. Nevertheless, (26) is a widely used model in many practical situations as it serves well as a motivating example to give a good sense as to what happens in more realistic channels.

Let us emphasize that there are many existing approaches in the literature to perform signal denoising, which can be roughly divided into two main categories: 1) denoising in the original signal domain (e.g., in time

or space [30]) and 2) denoising in the transform domain (e.g., using Fourier or a wavelet transform [31, 32]). Since the wavelet transform provides information in both the time and frequency domain, and the information in one scale is not contained in another scale, approaches in this second category can strike a balance between noise suppression and signal structural preservation, and have therefore been used widely in signal denoising. It is interesting to note that, usually, the detail coefficients of a noiseless signal are sparse. This means that in the wavelet transform most of the detail coefficients of a noiseless signal are very small/close to zero (as can be seen, for instance, in Fig. 4). So, the detail coefficients with a small magnitude can be considered as a noise component and can be set to zero. This is the basic idea of wavelet thresholding approaches, which are employed in signal denoising where the coefficients are compared with a threshold to determine whether they contain only noise or not.

It should be noted that since the approximation coefficients contain the low-frequency/important information of the signal, the thresholding is usually applied only to the detail coefficients (high-frequency components). The thresholding methods retain the significant components by setting to zero only the detail coefficients whose absolute values are less than a certain threshold. Most of the thresholding approaches try to find the optimal threshold value. The SureShrink method [33], proposed by Donoho and Johnstone, is an effective wavelet thresholding method for signal denoising that fits a level-dependent threshold to the wavelet coefficient. To show the performance of MSICA with respect to the other wavelet transforms, we extract the first and second-level detail coefficients of the noisy signal and use the SureShrink method for signal denoising. In our experiments, we also employ the FastICA method [34] and use the different standard test signals given in [35], i.e., Piece-Regular, Blocks, Doppler, and HeaviSine.

We consider now two scenarios for signal transmission and compare the signal denoising performance of MSICA in single and multi-channel transmissions.

Single-channel Transmission: In the first scenario, i.e., in the case of single-channel transmission as described in (26), if we divide the noisy signal into the even and odd samples, we obtain the vector of noisy observations as,

$$\mathbf{r} = \begin{bmatrix} r_1(n) \\ r_2(n) \end{bmatrix} = \begin{bmatrix} r(2n) \\ r(2n-1) \end{bmatrix} = \begin{bmatrix} x(2n) + w(2n) \\ x(2n-1) + w(2n-1) \end{bmatrix}, \quad (28)$$

where $w(2n)$ and $w(2n-1)$ have the same variance, σ_w^2 . The covariance matrix of \mathbf{r} can be written as,

$$\mathbf{C}_r = E\{\mathbf{r}\mathbf{r}^T\} = \sigma_r^2 \begin{bmatrix} 1 & \rho' \\ \rho' & 1 \end{bmatrix}, \quad (29)$$

where,

$$\sigma_r^2 = \sigma_x^2 + \sigma_w^2, \quad \rho' = \frac{\rho\sigma_x^2}{\sigma_x^2 + \sigma_w^2}. \quad (30)$$

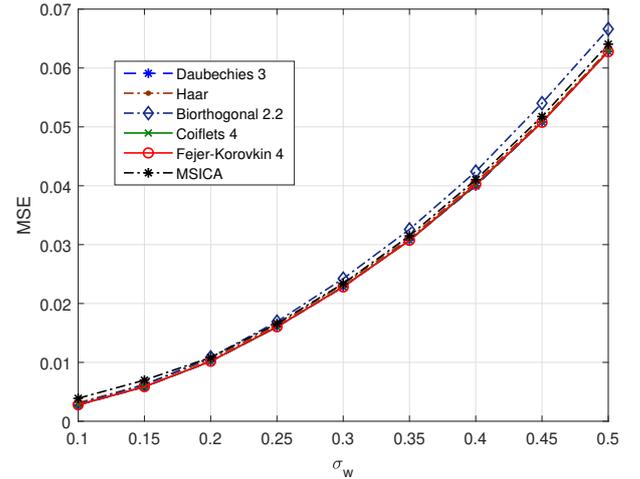


Fig. 6 – MSICA Mean Square Error (MSE) compared to well-known wavelet transforms.

Comparing (29) and (20), it is clear that the whitening matrix for the vector of noisy observations \mathbf{r} is like the one in (22),

$$\mathbf{W} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}. \quad (31)$$

Fig. 6 shows the performance of MSICA in the first scenario in terms of MSE against well-known wavelet transforms. As expected, MSICA shows similar performance compared to the other wavelet transform in suppressing the noise level and enhancing the quality of the signal as it is able to decompose the signal into approximation and detail. However, in the following we will explain how MSICA can have extraordinary performance when the odd and even samples of the original signal are passed through different channels.

Multi-channel Transmission: Let us consider now the second scenario where we investigate the performance of MSICA in a multi-channel transmission. Assume that two AWGN channels, named CH1 and CH2, are available to transmit the signal where, for example, we assume the variance of the noise in CH1 to be smaller than in CH2, which means that CH1 is more reliable and has better quality than CH2. In this case, if we transmit the even and odd samples through different channels, see Fig. 7(a), no matter through which one, then MSICA shows an extraordinary performance as it is able to filter out the noise of a noisier channel. In such a scenario, the even and odd samples of the noisy signal are,

$$r_1(n) = x(2n) + w_1(n), \quad (32a)$$

$$r_2(n) = x(2n-1) + w_2(n), \quad (32b)$$

where $w_1(n)$ and $w_2(n)$ are the AWGN in CH1 and CH2, respectively, and $\sigma_{w_2}^2 = K\sigma_{w_1}^2$ ($K > 1$). In this case, the covariance matrix of the vector of noisy observations \mathbf{r}

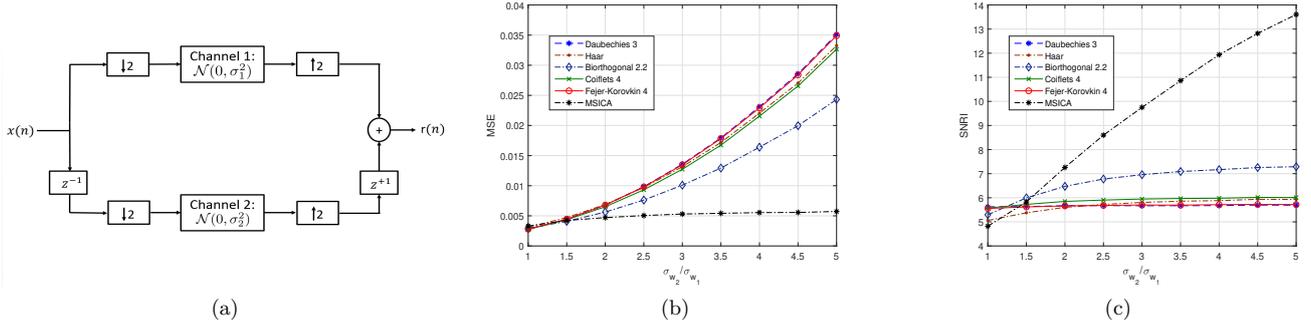


Fig. 7 – (a) Proposed multi-channel transmission where the even and odd samples of the original signal are transmitted through different channels and the receiver reconstructs the signal from the two outputs (z^{-1} and z^{+1} indicates time shift by $n=1$ to the right and left, respectively); (b) Mean Squared Error (MSE) vs. $\sigma_{w_2}/\sigma_{w_1}$ ($\sigma_{w_1} = 0.1$); (c) Performance of MSICA in terms of Signal-to-Noise Ratio Improvement (SNRI).

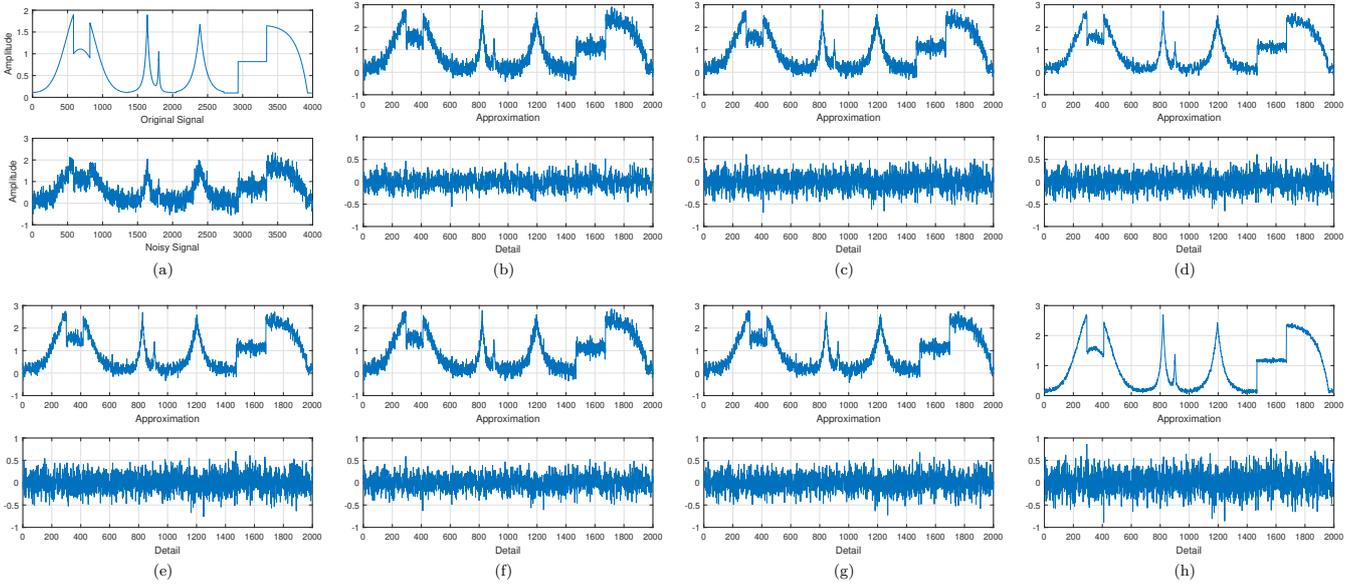


Fig. 8 – Comparison of MSICA with different wavelet transforms in decomposing a noisy PieceRegular signal. (a) Original Signal; (b) Approximation and detail by Daubechies 3 wavelet; (c) Approximation and detail by Haar wavelet; (d) Approximation and detail by Biorthogonal 2.2 wavelet; (e) Approximation and detail by Coiflets 4 wavelet; (f) Approximation and detail by Fejer-Korovkin 4 wavelet; (g) Approximation and detail by discrete Meyer wavelet; (h) Approximation and detail by MSICA.

is,

$$\mathbf{C}_r = \sigma_r^2 \begin{bmatrix} 1 & \rho' \\ \rho' & 1 + \frac{\rho' \sigma_{w_1}^2}{\sigma_r^2} \end{bmatrix}, \quad (33)$$

where,

$$\sigma_r^2 = \sigma_x^2 + \sigma_{w_1}^2, \quad \rho' = \frac{\rho \sigma_x^2}{\sigma_x^2 + \sigma_{w_1}^2}. \quad (34)$$

As it is clear from (33), in the case that the variance of the noise in CH1 and CH2 are different, the eigenvalues and eigenvectors of \mathbf{C}_r are dependent to the parameter K . This means that the low-pass and high-pass filters in the whitening process will be adaptive to the parameter K . In the following we will show that this adaptive filter is able to reject the effect of CH2 almost entirely. Figure 7(b) shows the performance of MSICA with respect to the other wavelet transforms when the original signal is passed through two different channels. As shown in Fig. 7(b), MSICA performance does not depend on the

variance of CH2, which means that MSICA is able to reject the AWGN of the CH2 from the noisy signal.

Moreover, to evaluate better the noise suppression performance, we have also examined the performance of MSICA in terms of Signal-to-Noise Ratio Improvement (SNRI),

$$\text{SNRI} = \text{SNR}_{out} - \text{SNR}_{in} = 10 \log \left(\frac{\sum_{n=1}^N (r(n) - x(n))^2}{\sum_{n=1}^N (\hat{x}(n) - x(n))^2} \right), \quad (35)$$

where SNR_{out} and SNR_{in} are the SNR of the denoised signal (output) and of the noisy signal (input), respectively. As shown in Fig. 7(c), the wavelet transforms have almost a fixed SNRI, whereas MSICA shows higher SNRI when the CH2 becomes noisier. This is because in MSICA the approximation and detail are statisti-

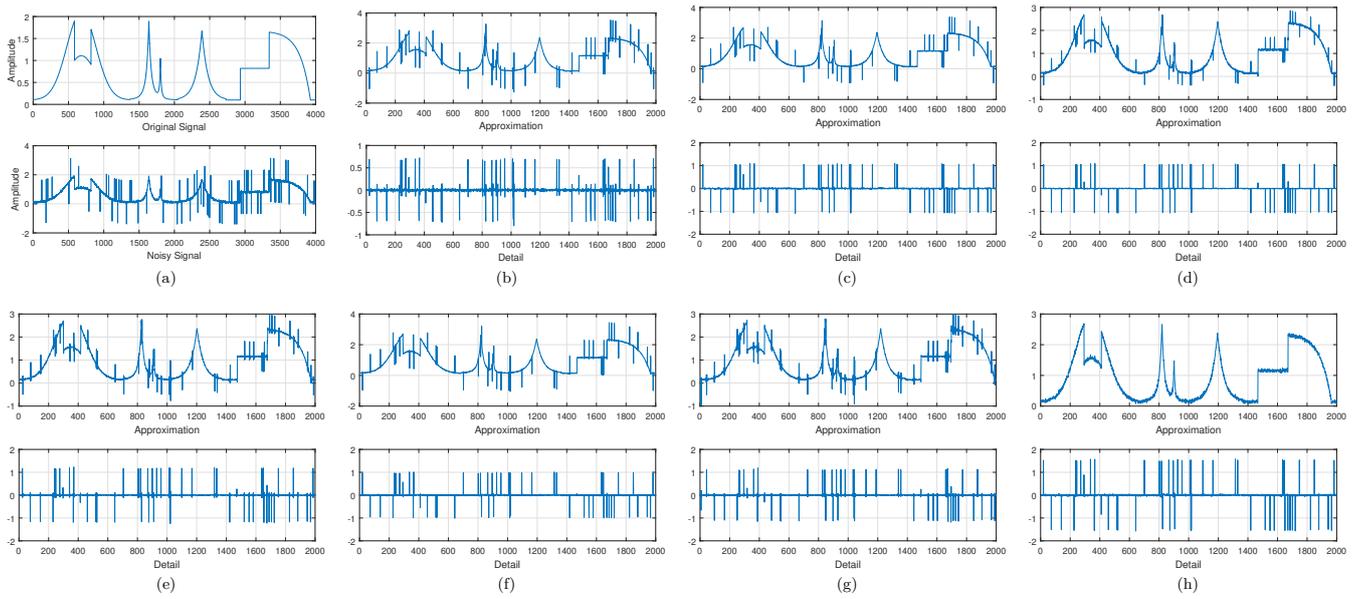


Fig. 9 – Comparison of MSICA with different wavelet transforms in decomposing a PieceRegular signal corrupted by impulse noise. (a) Original and Noisy Signal; (b) Approximation and detail by Daubechies 3 wavelet; (c) Approximation and detail by Haar wavelet; (d) Approximation and detail by Biorthogonal 2.2 wavelet; (e) Approximation and detail by Coiflets 4 wavelet; (f) Approximation and detail by Fejer-Korovkin 4 wavelet; (g) Approximation and detail by discrete Meyer wavelet; (h) Approximation and detail by MSICA.

cally independent; hence, MSICA is able to extract the noise signal from CH2 (via channel diversity), while the wavelet transforms are not able to do so.

Fig. 8 shows a signal decomposition where $\sigma_{w_2} = 0.2$, $\sigma_{w_1} = 0.05$. As it can be seen, the approximation obtained using MSICA is less noisy than the one obtained using the other wavelet transforms (Daubechies 3, Haar, Biorthogonal 2.2, Coiflets 4, Fejer-Korovkin 4, and Meyer). This result confirms our statement and shows that, because of the statistical independence between the approximation and detail, MSICA is able to extract the AWGN from the noisier channel.

In the other experiment, in order to show visibly that MSICA is able to extract the noise of CH2, we have explored its performance when the odd samples, passed through CH2, are corrupted by impulse noise. The Probability Density Function (PDF) of the impulse noise is given as,

$$P(w) = \begin{cases} P_a & w = a, \\ P_a & w = -a, \\ 1 - 2P_a & w = 0, \end{cases} \quad (36)$$

where $2P_a$ is the probability of existence of impulse noise in the received samples. In Fig. 9(a), the noisy signal is obtained by passing the even samples of the original signal through CH1 with AWGN with zero mean and $\sigma_{w_2}^2 = 0.004$, while the odd samples were passed through CH2 with impulse noise ($P_a = 0.01$ and $a = 1.5$). Fig. 9(b)-(h) show the performance of MSICA compared to a number of well-known wavelet transforms. As it is clear from Fig. 9(b)-(g), the traditional wavelet transforms (i.e., Daubechies 3, Haar, Biorthogonal 2.2, Coiflets 4, Fejer-Korovkin 4, Meyer) are not able to ex-

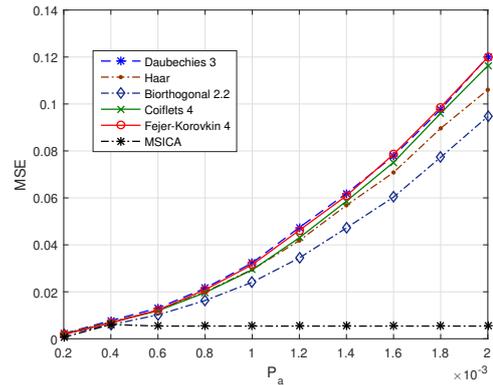


Fig. 10 – Impulse noise rejection in terms of Minimum Square Error (MSE); MSICA performance does not depend on P_a , whereas the performance of the other transforms decreases when P_a increases.

tract accurately the impulse components from the noisy signal. However, as it is shown in Fig. 9(h), MSICA is successful as the detail contains all the impulse components. This is because in MSICA the approximation and detail are statistically independent and, since the impulse noise is statistically independent from the original signal, MSICA can extract it in the detail coefficients.

Fig. 10 shows the performance of MSICA compared with different wavelet transforms when various values of P_a , as in (36), are considered. Here, the detail coefficients obtained by different methods have been set to zero to denoise the noisy signal. Since MSICA is able to extract the impulse noise, its performance does not depend on P_a , whereas the performance of the other transforms decreases when P_a increases.

To show that MSICA works on real signals too, we

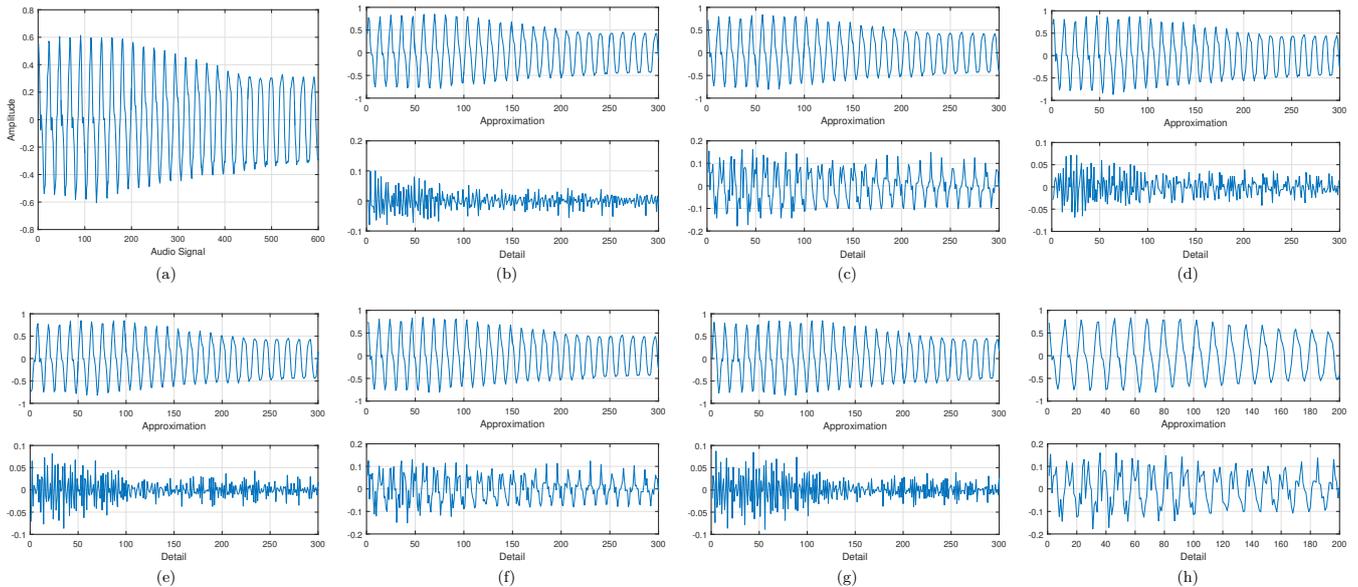


Fig. 11 – Approximation and detail coefficients obtained from an audio signal by some well-known wavelet transforms and MSICA. (a) Original audio Signal. Approximation and detail by (b) Daubechies 3; (c) Haar; (d) Biorthogonal 2.2; (e) Coiflets 4; (f) Fejer-Korovkin 4; (g) discrete Meyer; (h) MSICA.

have examined its performance on the signal depicted in Fig. 11(a), which is an audio signal with a sampling frequency equal to 16 KHz. Fig. 11(b)-(g) show the performance of the considered wavelet transforms in decomposing this signal into approximation and detail. As it is clear from Fig. 11(h), like the other transforms, MSICA is also able to decompose this audio signal into approximation and detail.

5. CONCLUSIONS

We presented MSICA, a new method for Multi-Scale decomposition based on Independent Component Analysis (ICA), where the approximation and detail are statistically independent. First, we extracted two correlated signals from the original digital signal by separating their even and odd samples; then, we proved that extracting the independent components of the correlated signals leads to the decomposition of the original signal into the approximation and detail. We showed that MSICA outperforms well-known wavelet transforms in signal denoising when transmitting a signal over multiple (noisy) channels as it exploits channel diversity. This property makes MSICA useful in many critical scenarios such as transmitting multimedia content reliably through underwater acoustic channels. Since these channels may vary quickly over time, it is difficult to estimate them, which makes transmitting multimedia content underwater a very challenging task.

REFERENCES

- [1] B. Ophir, M. Lustig, and M. Elad, “Multi-scale dictionary learning using wavelets,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1014–1024, 2011.
- [2] Y. Kopsinis and S. McLaughlin, “Development of emd-based denoising methods inspired by wavelet thresholding,” *IEEE Transactions on Signal Processing*, vol. 57, no. 4, pp. 1351–1362, 2009.
- [3] A. F. Laine, S. Schuler, J. Fan, and W. Huda, “Mammographic feature enhancement by multiscale analysis,” *IEEE Transactions on Medical Imaging*, vol. 13, no. 4, pp. 725–740, 1994.
- [4] F. M. Bayer, A. J. Kozakevicius, and R. J. Cintra, “An iterative wavelet threshold for signal denoising,” *Signal Processing*, vol. 162, pp. 10–20, 2019.
- [5] P. G. Bascosy, P. Quesada-Barriuso, D. B. Heras, and F. Argüello, “Wavelet-based multicomponent denoising profile for the classification of hyperspectral images,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 2, pp. 722–733, 2019.
- [6] A. Wong and A. K. Mishra, “Generalized probabilistic scale space for image restoration,” *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2774–2780, 2010.
- [7] G. Gilboa, “Nonlinear scale space with spatially varying stopping time,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2175–2187, 2008.
- [8] A. Mishra, A. Wong, D. A. Clausi, and P. W. Fieguth, “Quasi-random nonlinear scale space,” *Pattern Recognition Letters*, vol. 31, no. 13, pp. 1850–1859, 2010.
- [9] T. Lindeberg, “Scale-space theory: A basic tool for analyzing structures at different scales,” *Journal of applied statistics*, vol. 21, no. 1-2, pp. 225–270, 1994.
- [10] F. Jager, I. Koren, and L. Gyergyek, “Multiresolution representation and analysis of eeg waveforms,” in *Proceedings Computers in Cardiology*. IEEE, 1990, pp. 547–550.

- [11] P. Flandrin, G. Rilling, and P. Goncalves, "Empirical mode decomposition as a filter bank," *IEEE Signal Processing Letters*, vol. 11, no. 2, pp. 112–114, 2004.
- [12] B. Weng, M. Blanco-Velasco, and K. E. Barner, "ECG denoising based on the empirical mode decomposition," in *International Conference of the Engineering in Medicine and Biology Society*. IEEE, 2006, pp. 1–4.
- [13] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Computers in biology and medicine*, vol. 38, no. 1, pp. 1–13, 2008.
- [14] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 674–693, 1989.
- [15] S. Sardy, P. Tseng, and A. Bruce, "Robust wavelet denoising," *IEEE Transactions on Signal Processing*, vol. 49, no. 6, pp. 1146–1152, 2001.
- [16] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.
- [17] D. L. Donoho and J. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *biometrika*, vol. 81, no. 3, pp. 425–455, 1994.
- [18] G. Chen, T. D. Bui, and A. Krzyżak, "Image denoising with neighbour dependency and customized wavelet and threshold," *Pattern Recognition*, vol. 38, no. 1, pp. 115–124, 2005.
- [19] A. Witkin, "Scale-space filtering: A new approach to multi-scale description," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 9. IEEE, 1984, pp. 150–153.
- [20] A. Wong and X. Y. Wang, "A bayesian residual transform for signal processing," *IEEE Access*, vol. 3, pp. 709–717, 2015.
- [21] B. Mijovic, M. De Vos, I. Gligorijevic, J. Taelman, and S. Van Huffel, "Source separation from single-channel recordings by combining empirical-mode decomposition and independent component analysis," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 9, pp. 2188–2196, 2010.
- [22] S. Wang and C. J. James, "On the independent component analysis of evoked potentials through single or few recording channels," in *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2007, pp. 5433–5436.
- [23] M. E. Davies and C. J. James, "Source separation using single channel ica," *Signal Processing*, vol. 87, no. 8, pp. 1819–1832, 2007.
- [24] G.-J. Jang, T.-W. Lee, and Y.-H. Oh, "Single-channel signal separation using time-domain basis functions," *IEEE Signal Processing Letters*, vol. 10, no. 6, pp. 168–171, 2003.
- [25] J. Miettinen, M. Matilainen, K. Nordhausen, and S. Taskinen, "Extracting conditionally heteroskedastic components using independent component analysis," *Journal of Time Series Analysis*, vol. 41, no. 2, pp. 293–311, 2020.
- [26] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.
- [27] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, "A review of blind source separation methods: two converging routes to ilarma originating from ica and nmf," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019.
- [28] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4, pp. 411–430, 2000.
- [29] P. Comon, "Independent component analysis, a new concept?" *Transactions on Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [30] I. Kauppinen and K. Roth, "Improved noise reduction in audio signals using spectral resolution enhancement with time-domain signal extrapolation," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 6, pp. 1210–1216, 2005.
- [31] S. Durand and J. Froment, "Artifact free signal denoising with wavelets," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 6, 2001, pp. 3685–3688.
- [32] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, A. K. Abasi, and S. N. Makhadmeh, "Eeg signals denoising using optimal wavelet transform hybridized with efficient metaheuristic methods," *IEEE Access*, vol. 8, pp. 10 584–10 605, 2019.
- [33] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [34] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
- [35] R. R. Coifman and D. L. Donoho, *Translation invariant denoising*. Springer, 1995.

AUTHORS



Abolfazl Hajisami received his PhD degree in Electrical and Computer Engineering (ECE) from Rutgers University, NJ, USA, in 2018. He received his MSc and BSc degrees from Sharif University of Technology and from Shahid Beheshti University (Tehran, Iran), in 2010 and 2008, respectively. His research interest lies in wireless communication, vehicle-to-everything (V2X) communication, and signal processing. His PhD thesis dealt with spectral and energy efficiency in cloud radio access networks. He is now a Staff Research Engineer at Qualcomm, where he works on V2X technologies.



Dario Pompili is an Assoc. Prof. with the Dept. of ECE at Rutgers, where he directs the Cyber-Physical Systems Laboratory (CPS Lab), which focuses on mobile edge computing, wireless networking, acoustic communications, and sensor networks. He received his PhD in ECE from the Georgia Institute of Technology in 2007. He had previously received his ‘Laurea’ (combined BS and MS) and Doctorate degrees in Telecommunications and System Engineering from the U. of Rome “La Sapienza,” Italy, in 2001 and 2004, respectively. He has received a number of awards including the NSF CAREER’11, ONR Young Investigator Program’12, and DARPA Young Faculty’12 awards. In 2015, he was nominated Rutgers-New Brunswick Chancellor’s Scholar. He has served in various roles on many international conferences. He published about 150 refereed scholar publications, some of which selected to receive best paper awards: with about 11,600 citations, Dr. Pompili has an h-index of 42 and an i10-index of 101 (Google Scholar, Oct’20). He is a Senior Member of the IEEE Communications Society (2014) and a Distinguished Member of the ACM (2019). He is an Area Editor for IEEE Transactions on Mobile Computing (TMC).