Analysis of the Reconstruction of Sparse Signals in the DCT Domain Applied to Audio Signals

Ljubiša Stanković, *Fellow, IEEE,* and Miloš Brajović, *Student Member, IEEE*

***Abstract*—Sparse signals can be reconstructed from a reduced set of signal samples using compressive sensing (CS) methods. The discrete cosine transform (DCT) can provide highly concentrated representations of audio signals. This property implies the DCT as a good sparsity domain for the audio signals. In this paper, the DCT is studied within the context of sparse audio signal processing using the CS theory and methods. The DCT coefficients of a sparse signal, calculated with a reduced set of available samples, can be modeled as random variables. It has been shown that the statistical properties of these variables are closely related to the unique reconstruction conditions. The main result of the paper is in an exact formula for the mean square reconstruction error in the case of approximately sparse and nonsparse noisy signals, reconstructed under the sparsity assumption. Based on the presented analysis, a simple and computationally efficient reconstruction algorithm is proposed. The presented theoretical concepts and the efficiency of the reconstruction algorithm are verified numerically, including examples with synthetic and recorded audio signals with unavailable or corrupted samples. Random disturbances and disturbances simulating clicks or inpainting in audio signals are considered. Statistical verification is done on a dataset with experimental signals. Results are compared with some classical and recent methods used in similar signal and disturbance scenarios.**

***Index Terms*—Audio signals, digital signal processing, compressed sensing, discrete cosine transform, sparse signal processing.**

I. Introduction

SPARSE signals are characterized by a small number of nonzero coefficients in one of their transformation domains [1]–[22]. These signals can be reconstructed from a reduced set of measurements [1]–[15], [21]. Measurements are linear combinations of the sparsity domain coefficients. Signal samples can be considered as measurements in the case of linear signal transforms. In certain applications reduced sets of measurements/samples result as a consequence of their physical unavailability, whereas in other applications they are a result of a particular interest to reduce the number of measurements while preserving the whole information (data compression) [1], [2]. The unavailability of signal samples may also arise in the cases when some samples are intentionally omitted due to a high noise or corruption [4]. The last scenario may happen in highly corrupted audio signals.

In signal processing, the most common transformation domain is the Fourier domain [4], [14], with the discrete-time domain of signal samples as the measurements [1], [2]. Corresponding measurement matrices are the partial Discrete Fourier transform (DFT) matrix and the partial random Fourier transform matrix [1], [2]. The influence of a reduced set of samples on the analysis and signal reconstruction/synthesis with the partial Fourier transform matrices is studied in [4].

The discrete cosine transform (DCT) is one important and commonly used tool in audio signal processing [14]. The main reason is that the audio signals can be represented in a more compact form in the DCT domain than in the Fourier domain. This is the reason why many signal compression algorithms exploit the DCT [14], [23]–[25]. This particular transform was also used in speech enhancement applications based on compressive sensing due to its superior compressibility [7]. Therefore, this transform can play an important role in the audio signal representation with a reduced set of samples. The measurement matrices obtained from the DCT transform matrices are the topic of this paper. Like in the case of the DCT itself, many specific properties of DCT partial measurement matrices make their analysis different from the analysis of the Fourier transform based partial matrices.

The initial idea for this analysis comes from our previous correspondence on the two-dimensional DFT and radar signals [15]. However, the DCT sparsity domain exhibits many properties different from the two-dimensional DFT, starting from the fact that the ℓ_2-norms of the partial DCT matrix columns are random variables. The ℓ_2-norms of the partial DFT matrix columns are constant.

Commonly, audio signals are subject to localized time-domain distortions, including impulsive noises and clicks [26]–[41], clipping [26], packet loss during the signal transmission [42]–[50], and CD scratching [26], [27]. Significant research efforts have been focused on the removal and reconstruction/synthesis of the audio signals with this kind of disturbances [27], [35], [37]–[41], [51]–[55]. Corruption of audio signals by clicks in old recordings, scratched CDs, or the typed keystrokes [51], [56], assumes corrupted samples or intervals of corrupted samples occurring at random locations. Many different approaches have been proposed to recover the corrupted samples, including the median and low-pass filtering, autoregressive modeling [37], [38], and the Bayesian estimation [41]. Recently, the emerging area of CS provided new approaches for restoration of corrupted samples [7], [8], [26], [52]. Spectrogram as a domain of sparsity, in conjunction with solving a regularized ℓ_1-norm least squares problem, has been considered in [52] for treating a speech recognition problem. It has been shown that the problem of corrupted/unavailable samples can be efficiently solved using the matching pursuit (MP) approaches [8], [26] with the DCT acting as a domain of the audio signal sparsity. Audio inpainting concept, introduced

Authors are with the Faculty of Electrical Engineering, University of Montenegro, Cetinjski put b.b., 81000 Podgorica, Montenegro, (e-mail: ljubisa@ac.me, milosb@ac.me)

in [26], assumes the reconstruction of audio signal portions distorted by disturbances such as impulsive noise, clicks, or audio clipping, using the orthogonal MP algorithms. A variant of the reconstruction algorithms from this class, particularly adapted to the time-domain noise (Tdn-CoSaMP), has been recently introduced in [8] for speech enhancement. Therein, a random sampling matrix is included in the sensing scheme. Substantial efforts have been made to derive the reconstruction error upper bounds for this CS reconstruction algorithm.

In the CS theory, only upper bounds of the mean square reconstruction error are derived [3]. The upper bounds introduced in [8] are similar to these bounds in the general CS theory. The main contribution of this paper is in the exact relation for the mean squared error (MSE) in audio signals reconstructed from a reduced set of signal samples when the DCT is used as the sparsity domain.

The exact reconstruction error relation is the final result of a comprehensive analysis of the influence of unavailable/corrupted samples to the DCT, presented in the paper. This analysis shed a new light on some other important relations in the CS theory. A simple derivation of the coherence-based condition for unique reconstruction is presented and explained for the partial DCT measurement matrices. The presented reconstruction error analysis has also resulted in a simple and computationally efficient method for sparse signal reconstruction, with a data-driven threshold. This algorithm belongs to the class of MP algorithms. The results for additive noise influence are derived and related to the results obtained by using the Bayesian-based approach to the reconstruction of noisy signals sparse in the DCT domain [12]. The analysis of additive noise influence is combined with the derived initial CS noise properties, to get the main result that consists in the exact expression for the MSE in the reconstructed signal. The presented theory is illustrated and verified numerically on various audio signals.

The paper is organized as follows. In Section II the basic DCT definitions are given. Starting from the reduced set of observations framework and the partial DCT matrix, in Section III we present the theorem describing statistical properties of the DCT coefficients of randomly under-sampled data, the reconstruction based on missing samples analysis as well as the coherence reconstruction relation. In Section IV additive noise influence on the reconstruction result is analyzed, whereas the nonsparse signal reconstruction scenario is analyzed in Section V. The application of the presented theory to audio signals is illustrated in Section VI. Validation of the theory in the audio signal processing context is done in Section VII.

II. Basic definitions

The DCT (DCT-II) of a discrete-time signal $x(n)$ is defined by

$$X^C(k) = \sum_{n=0}^{N-1} a_k x(n) \cos\left(\frac{\pi(2n+1)}{2N}k\right), \qquad (1)$$

with $k = 0, ..., N-1$, while the corresponding inverse transform has the form

$$x(n) = \sum_{k=0}^{N-1} a_k X^C(k) \cos\left(\frac{\pi(2n+1)}{2N}k\right), \qquad (2)$$

$n = 0, ..., N-1$, where $a_k = \sqrt{1/N}$ for $k = 0$ and $a_k = \sqrt{2/N}$ for $k \neq 0$. The DCT transform can be written in a matrix form

$$\mathbf{X}^C = (\mathbf{C}_N)\mathbf{x}, \qquad (3)$$

where $\mathbf{X}^C$, $(\mathbf{C}_N)$, and $\mathbf{x}$ are the DCT coefficients vector, DCT transformation matrix and the signal vector, respectively. For the inverse DCT the relation $\mathbf{x} = (\mathbf{C}_N)^{-1}\mathbf{X}^C$ holds. Note that for this DCT matrix relation $(\mathbf{C}_N)^{-1} = (\mathbf{C}_N)^T$ holds [57]. Since we will use the DCT-II form in this paper, index II will be omitted.

A signal of the form:

$$x(n) = \sum_{l=1}^{K} a_{k_l} A_l \cos\left(\frac{\pi(2n+1)}{2N}k_l\right) \qquad (4)$$

is sparse in the DCT domain if the number of components (nonzero DCT coefficients) K is much smaller than the number of signal samples N, $K \ll N$. Component amplitudes are denoted by $A_l,\ l = 1, 2, \ldots, K$. Positions $k_1, k_2, \ldots, k_K$ will be referred to as signal coefficient positions while the remaining positions will be referred to as nonsignal coefficient positions.

The DCT of signal (4) reads:

$$X^C(k) = \sum_{n=0}^{N-1} \sum_{l=1}^{K} A_l a_k a_{k_l} \cos\left(\tfrac{\pi(2n+1)}{2N}k_l\right) \cos\left(\tfrac{\pi(2n+1)}{2N}k\right) \qquad (5)$$

where $k = 0, ..., N-1$. Signal components of the form $A_l \cos(\pi k_l(2n+1)/(2N))$ are multiplied with DCT basis functions, producing in (5) the terms of the form:

$$z(k_l, k, n) = A_l a_k a_{k_l} \cos\left(\tfrac{\pi(2n+1)}{2N}k_l\right) \cos\left(\tfrac{\pi(2n+1)}{2N}k\right). \qquad (6)$$

If all signal samples are available, then the corresponding DCT equals $X^C(k) = \sum_{l=1}^{K} A_l \delta(k - k_l)$.

III. Reduced set of observations

Assume that only $M \leq N$ randomly positioned signal samples at $n_i \in \mathbf{M} = \{n_1, n_2, ..., n_M\} \subseteq \mathbf{N} = \{0, 1, ..., N-1\}$ are available,

$$\mathbf{y} = \{x(n_1), x(n_2), \ldots, x(n_M)\} \subseteq \mathbf{x}$$

with

$$x(n_i) = \sum_{k=0}^{N-1} a_k X^C(k) \cos\left(\frac{\pi(2n_i+1)}{2N}k\right),\ i = 1, ..., M.$$

A matrix form of the available samples is

$$\mathbf{y} = \mathbf{A}_{MN}\mathbf{X}^C,$$

with $\mathbf{A}_{MN}$ representing an $M \times N$ matrix of observations (measurements matrix). It is defined as the partial inverse DCT matrix, with rows being equal to the rows of $(\mathbf{C}_N^{-1})$, corresponding to the available samples positions n_i:

$$\mathbf{A}_{MN} = \frac{\sqrt{2}}{\sqrt{N}} \begin{bmatrix} \frac{\sqrt{2}}{2} & \cos\left(\frac{\pi(2n_1+1)}{2N}\right) & \cdots & \cos\left(\frac{\pi(2n_1+1)(N-1)}{2N}\right) \\ \frac{\sqrt{2}}{2} & \cos\left(\frac{\pi(2n_2+1)}{2N}\right) & \cdots & \cos\left(\frac{\pi(2n_2+1)(N-1)}{2N}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\sqrt{2}}{2} & \cos\left(\frac{\pi(2n_M+1)}{2N}\right) & \cdots & \cos\left(\frac{\pi(2n_M+1)(N-1)}{2N}\right) \end{bmatrix}.$$

In compressive sensing, it is common to normalize the column mean value energies (diagonal elements of matrix $\mathbf{A}_{MN}^T\mathbf{A}_{MN}$). In this case, the factor $\sqrt{M/2}$ would be used instead of $\sqrt{N/2}$.

The initial (norm-two based) DCT estimation uses the available samples only

$$X_0^C(k) = \sum_{n\in\mathbf{M}} a_k x(n)\cos\left(\frac{\pi(2n+1)}{2N}k\right) = \sum_{i=1}^{M}\sum_{l=1}^{K} z(k_l,k,n_i), \quad (7)$$

where $k = 0, 1, ..., N-1$. It produces the same result as if the missing (unavailable) samples assume zero values [4]. In a matrix form we can write

$$\mathbf{X}_0^C = \mathbf{A}_{MN}^T\mathbf{y}.$$

Note that the terms $z(k_l,k,n_i)$ belong to the set

$$\mathbf{\Theta} = \{z(k_l,k,n_1),\ z(k_l,k,n_2), ...,\ z(k_l,k,n_M)\},$$

that is a subset of complete set of samples

$$\Big\{A_l a_k a_{k_l}\cos\left(\frac{\pi(2n+1)}{2N}k_l\right)\cos\left(\frac{\pi(2n+1)}{2N}k\right), \quad n,\ k=0,...,N-1,\ l=1,...,K\Big\}.$$

Consequently, the set of missing samples $\mathbf{Q}$ can be considered as a subset of the complete set of samples, $\mathbf{Q} = \mathbf{N}\backslash\mathbf{M}$. Original signal samples at the positions defined by $\mathbf{Q}$ are affected by a noise [4]:

$$\eta(k_l,k,n) = \begin{cases} -z(k_l,k,n), & n\in\mathbf{Q} \\ 0, & n\in\mathbf{M}, \end{cases} \quad (8)$$

for $k = 0, ..., N-1,\ l = 1, ..., K$. The DCT coefficients

$$X_0^C(k) = \sum_{i=1}^{M}\sum_{l=1}^{K} z(k_l,k,n_i) = \sum_{n=0}^{N-1}\sum_{l=1}^{K}[z(k_l,k,n)+\eta(k,k_l,n)] \quad (9)$$

with randomly positioned available samples and $1 \ll M \ll N$ may be considered as random variables. Here we will analyze the statistical properties of coefficients $X_0^C(k)$.

Theorem 1: Assume a sparse signal with K nonzero coefficients in the DCT domain at random positions k_l with amplitudes A_l, $l = 1, 2, ..., K$. Assume that out of the total number of N samples only M samples, $1 \ll M \ll N$, are available. The DCT coefficients $X_0^C(k)$ calculated using the available samples are random, approximately Gaussian distributed variables. Their mean-value and variance are

$$\mu_{X_0^C(k)} = \frac{M}{N}\sum_{l=1}^{K} A_l\delta(k-k_l) \quad (10)$$

and

$$\sigma^2_{X_0^C(k)} = \frac{M(N-M)}{N^2(N-1)}\sum_{l=1}^{K}A_l^2[1-\frac{1}{2}\delta(k-(N-k_l)) - \frac{1}{2}[1+\delta(k_l)]\delta(k-k_l)], \quad (11)$$

respectively, where a_k are the DCT constants.

The proof is given in the Appendix A. The theorem result will be numerically checked next.

Example 1: A sparse monocomponent randomly undersampled signal of form (4) is considered, with $K = 1$, $k_1 = 12$, and $N = 129$. The number of randomly positioned available samples M is varied from 0 to $N-1$. For each number of available samples M, the variance of the DCT coefficient $X_0^C(k_1)$ at the signal component position is calculated and averaged in 30000 independent realizations with randomly positioned missing samples. The result is compared with the theoretical variance (11). The results are shown in Fig. 1(a).

The same experiment is performed for the DCT coefficients corresponding to the nonsignal positions $k \neq k_1$ (and for $k \neq N-k_1$). Since we now have $N-2 = 127$ non-signal (noise) coefficients in one realization, we reduce the set of random realizations to 200 (with the total number of observed nonsignal coefficients in all realizations being 25400). The average variance is compared with the theoretical result (11) in Fig. 1(b). To emphasize the difference between variances at the signal coefficient and the nonsignal coefficients positions, the variances of all DCT coefficients of the signal (for $M = 64$ available samples) are calculated based on 10000 independent realizations and shown in Fig. 1(c). Finally, the signal component position k_1 is varied from 0 to $N-1$ with $M = 50$ available samples. Based on 10000 independent realizations of randomly positioned missing samples, the variance of $X_0^C(k_1)$ is calculated and compared with the theoretical result (11). As expected, a position independent variance value is obtained, except for $k_1 = 0$ when the variance is zero (as expected). The results are shown in Fig. 1(d).

Example 2: A multicomponent signal of form (4), with $K = 5$, is considered here. The complete signal length is $N = 155$. Signal components are positioned at $k_l = \{22, 49, 47, 89, 100\}$, with corresponding amplitudes $A_l = \{5, 3.5, 1.5, 2.5, 1\}$. In order to check the distribution of random variables $X_0^C(k)$, 60000 independent realizations of signal with randomly positioned $M = 90$ samples are considered. The histograms of the signal component coefficient $X_0^C(22)$ and the nonsignal coefficient $X_0^C(130)$ are shown in Fig. 2. Histograms are scaled with the number of realizations. The histogram of coefficient $C(22)$ is compared with the Gaussian distribution (in dots) having the mean value $\mu_{X_0^C(22)} = A_1 M/N \approx 2.90$ and the variance $\sigma^2_{X_0^C(22)} = 0.0542$, calculated according to (11). The result is presented in Fig. 2 (right). Similarly, the nonsignal coefficient $X_0^C(130)$ histogram is compared with the zero-mean Gaussian distribution, whose variance is equal to $\sigma^2_{X_0^C(130)} = 0.0739$, according to (11). The result is shown in Fig. 2 (left).

A. The Reconstruction Based on the Missing Sample Analysis

The variance of noise only components, using the result of Theorem (11) can be expressed as:

$$\sigma^2_{X_0^C(k)} = \frac{M(N-M)}{N^2(N-1)}\sum_{l=1}^{K}A_l^2\left[1-\frac{1}{2}\delta(k-(N-k_l))\right],$$

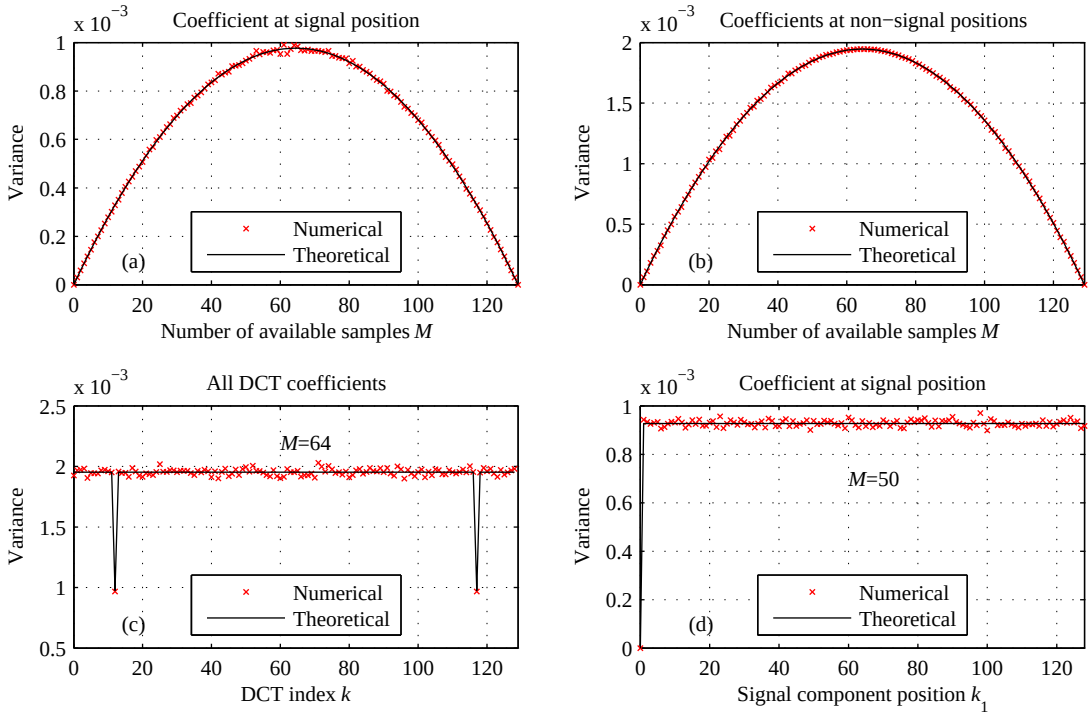

Figure 1. Numerical check of the derived variance: (a) the variance of the DCT coefficient at the signal position $k = k_1$ as a function of the number of available samples, (b) the average variance of the DCT coefficients at noise only positions $k \neq k_1$ shown as a function of the number of available samples, (c) the variance of all DCT coefficients of a sparse mono-component signal with $k_1 = 12$ and with $M = 64$ available samples, (d) the variance of the DCT coefficient $X_0^C(k = k_1)$ as a function of k_1, calculated for signals with $M = 50$ available samples.

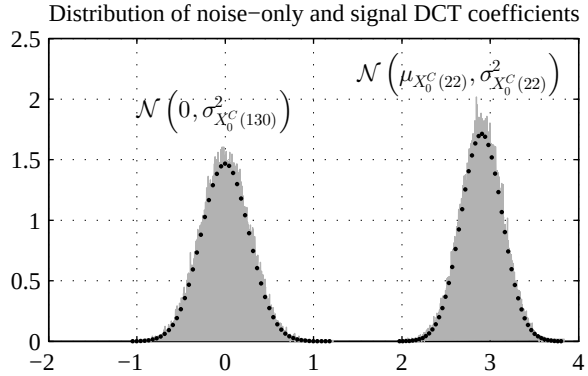

Figure 2. Scaled histograms of DCT coefficients and the corresponding pdfs: at the non-signal position (left) and at the signal position (right). Dots denote the theoretical result.

where $k \neq k_l$. The coefficients at positions $k = N - k_l$, $l = 1, ..., K$ have the variance reduced for $A_l^2/2$ when compared to the other coefficients at nonsignal positions. We may assume that the variance for noise-only coefficients is position independent and equal to:

$$\sigma^2_{csN} = \frac{M(N-M)}{N^2(N-1)} \sum_{l=1}^{K} A_l^2. \tag{12}$$

It is overestimated at positions $k = N - k_l$. The variance (12) depends on the total signal power, that can be easily estimated using the available samples, as

$$\sum_{l=1}^{K} A_l^2 = E_s \cong \frac{N}{M} \sum_{n \in \mathbf{M}} s^2(n).$$

If we set a threshold, for example at $4\sigma_{csN}$, then (according to the four-sigma rule) we know that less than 1 noise-only (nonsignal) coefficient in 15,000 coefficients will be above this level. Coefficients above the threshold may be considered as signal components and reconstructed after their positions are detected. Algorithms for this type of reconstruction are given in Appendix B. If a noise component is included, then it will produce a zero coefficient value in the final result. In the case when small signal coefficients exist, within the level of the noise produced by missing samples, the reconstruction procedure can be repeated after the strongest components are detected, reconstructed and subtracted from the available samples values.

Example 3: A three-component signal with $N = 256$ is observed. Component positions and amplitudes are $k_1 = 14$, $k_2 = 162$, $k_3 = 203$ and $A_1 = 1$, $A_2 = 1/\sqrt{2}$, $A_3 = 1/2$, respectively. Only $M = 128$ of its randomly positioned samples are available. In order to reconstruct the signal, we calculate the initial estimation $X_0^C(k)$. Then we define a threshold based on the variance due to the missing samples, whose value is $\sigma^2_{csN} = \frac{128(256-128)}{256^2(256-1)}(1+1/2+1/4) = 0.0017$. A threshold at $4\times\sqrt{0.0017} = 0.1657$ is used, Fig. 3. The weakest component mean value $1/4$ is well above this threshold.

This kind of analysis can be generalized to a K component signal. In the worst case, we should be able to detect at least

the strongest component (the detection of other components would follow after the reconstruction and subtraction of this strongest component). The worst case for the strongest component detection would be the case when other components are equally strong, $A_1 = A_2 = \cdots = A_K = 1$. In the worst case of a K-sparse signal with equal components, the four-sigma threshold would be

$$T = 4\sqrt{\frac{M(N-M)}{N^2(N-1)}K}.$$

We may conclude that the mean value of a signal-only coefficient is above the threshold (with a probability defined by the four-sigma rule) if

$$\frac{M}{N} > 4\sqrt{\frac{M(N-M)}{N^2(N-1)}K}$$

holds. It produces the upper bound for sparsity K

$$K < \frac{M(N-1)}{16(N-M)}. \tag{13}$$

For $N = 256$ and $M = 128$ we get $K < 15.94$ or $K \leq 15$. If we allowed the three-sigma rule, then $K \leq 28$ would be obtained. Note that few noise components, that are wrongly detected as the signal components, will not influence the reconstruction as far as the reconstruction conditions are met. The algorithm will produce zero values as the result for such coefficients.

After the component positions are detected, the measurement equation becomes

$$\mathbf{y} = \mathbf{A}_{MK}\mathbf{X}_K^C,$$

where $\mathbf{X}_K^C$ is the vector of K unknown coefficients at the detected positions. The measurement matrix is reduced to a $M \times K$ matrix by omitting columns corresponding to the zero-valued coefficients. Since $K < M$, the equation can be solved in the mean-square sense. The result is a vector with K elements

$$\mathbf{X}_K^C = (\mathbf{A}_{MK}^T\mathbf{A}_{MK})^{-1}\mathbf{A}_{MK}^T\mathbf{y}. \tag{14}$$

If the obtained coefficients are such that $\mathbf{e} = \mathbf{y} - \mathbf{A}_{MK}\mathbf{X}_K^C$ is zero (or within the acceptable bounds), then we have found the solution. If the error is not small, then some of non-zero coefficients have not been detected and included into $\mathbf{X}_K^C$. The calculation should be repeated with $\mathbf{e}$ now acting as a signal. The candidates for nonzero coefficient positions should be detected based on the initial DCT of this new signal and added to the previous set $\mathbf{K}$ of nonzero coefficient positions. Calculation of $\mathbf{X}_K^C$ should be repeated with this new (updated) set of positions, until an acceptable (zero) error level is achieved. The summary for this reconstruction procedure is given in Appendix B.

Next, the exact analysis of the signal-only and noise-only (non-signal) coefficients will be presented, in order to better position the threshold for the signal components detection.

Let us observe a K-sparse signal of the form (4). The noise-only DCT coefficient is a random variable described by approximative Gaussian distribution, $\mathcal{N}(0, \sigma_{csN}^2)$ with zero mean and variance σ_{csN}^2 defined by (12). The l-th signal DCT coefficient is also a random variable described by approximative Gaussian distribution $\mathcal{N}(\frac{M}{N}A_l, \sigma^2_{X_0^C(k_l)})$, $l = 1, 2, ..., K$, with mean value $\mu_{X_0^C(k_l)} = \frac{M}{N}A_l$ and variance $\sigma^2_{X_0^C(k_l)}$ defined by (11). The absolute DCT coefficient values at the position of the l-th signal component have the folded normal distribution

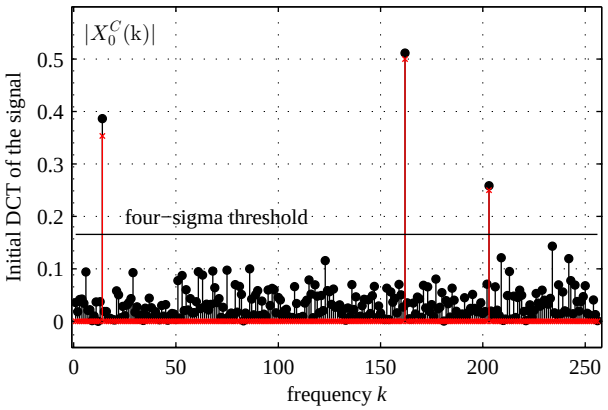

Figure 3. The DCT of a three-component sparse signal, with the four-sigma threshold (horizontal line): crosses - DCT coefficients of the full-length signal, dots - DCT coefficients of the signal with missing samples.

$$p(\xi) = \frac{1}{\sigma_{X_0^C}(k_l)\sqrt{2\pi}}\Big[\exp\left(-\frac{(\xi - \frac{M}{N}A_l)^2}{2\sigma^2_{X_0^C(k_l)}}\right) + \exp\left(-\frac{(\xi + \frac{M}{N}A_l)^2}{2\sigma^2_{X_0^C(k_l)}}\right)\Big]. \tag{15}$$

The probability density function for the absolute values of noise-only DCT coefficients is the half normal distribution

$$q(\xi) = \frac{\sqrt{2}}{\sigma_N\sqrt{\pi}}\exp\left(-\frac{\xi^2}{2\sigma_{csN}^2}\right).$$

The DCT coefficient at a noise-only position takes a value lower than Ξ, with probability

$$Q(\Xi) = \int_0^{\Xi}\frac{\sqrt{2}}{\sigma_N\sqrt{\pi}}\exp\left(-\frac{\xi^2}{2\sigma_{csN}^2}\right)d\xi = \mathrm{erf}\left(\frac{\Xi}{\sqrt{2}\sigma_{csN}}\right). \tag{16}$$

The total number of noise-only coefficients is $N - K$. The probability that $N - K$ independent DCT noise-only coefficients are lower than Ξ is $Q(\Xi)^{N-K}$. Probability that at least one of $N-K$ DCT noise-only coefficients is greater than Ξ is $G(\Xi) = 1 - Q(\Xi)^{N-K}$. When a noise-alone DCT value surpasses the DCT coefficient at a signal position, then an error in the signal component detection occurs. To calculate this error probability, consider the absolute DCT value of a signal component at and around ξ. The DCT coefficient at the signal position has a value within ξ and $\xi + d\xi$ with the probability $p(\xi)d\xi$, where $p(\xi)$ is defined by (15). The probability that at least one of $N - K$ DCT noise-alone coefficients is above ξ is $G(\xi) = 1 - Q(\xi)^{N-K}$. Consequently, the probability that the absolute value of a DCT signal-only coefficient is within

ξ and $\xi + d\xi$ and that at least one of the absolute DCT noise-alone values outside the DCT signal value exceeds the DCT signal value is $G(\xi)p(\xi)d\xi$. Considering all possible values of ξ, it follows that the probability of the wrong detection of the l-th signal component is

$$P_E = \frac{1}{\sigma_{X_0^C(k_l)}\sqrt{2\pi}} \int_0^\infty \left(1 - \text{erf}\left(\frac{\xi}{\sqrt{2}\sigma_{csN}}\right)^{N-K}\right) \times \left[\exp\left(-\frac{(\xi - \frac{M}{N}A_l)^2}{2\sigma^2_{X_0^C(k_l)}}\right) + \exp\left(-\frac{(\xi + \frac{M}{N}A_l)^2}{2\sigma^2_{X_0^C(k_l)}}\right)\right] d\xi.$$

A simple approximation of this expression can be obtained following the steps for the DFT analysis presented in [4].

B. Coherence Reconstruction Relation

For the worst case analysis, assume the maximal possible influence from other signal components to the strongest signal component detection. The maximal influence of signal components to each other occurs when all K components are with equal (unity) amplitudes.

For M available samples, the mean value of a component (DCT coefficient at a signal position) is M/N. The noise component at the frequency index k, that originates from the component at k_l, is equal to

$$Q(k,k_l) = \sum_{n\in\mathbf{M}, k\neq k_l} a_k a_{k_l} \cos\left(\frac{\pi(2n+1)}{2N}k_l\right) \times \cos\left(\frac{\pi(2n+1)}{2N}k\right). \quad (17)$$

It can be related to the coherence factor of $\mathbf{A}_{MN}^T\mathbf{A}_{MN}$, defined by

$$\mu = \max\Big|\frac{N}{M}\sum_{n\in\mathbf{M}, k\neq k_l} a_k a_{k_l} \cos\left(\frac{\pi(2n+1)}{2N}k_l\right) \times \cos\left(\frac{\pi(2n+1)}{2N}k\right)\Big|. \quad (18)$$

If the noise originating from all signal components is such that it adds up in phase at the nonsignal coefficient positioned at k, assuming maximal possible value for μ, then the maximal nonsignal coefficient value is

$$K\max|Q(k,k_l)| = K\mu\frac{M}{N}.$$

At the same time, if at a signal coefficient position all noise factors that originate from other $K-1$ components add up in phase in the negative direction from the component mean value, assuming again their maximal possible absolute values μ, then the resulting worst signal component amplitude will be

$$\min\{X_0^C(k_l)\} = \frac{M}{N} - (K-1)\max|Q(k,k_l)|.$$

The detection of the signal component is still possible if

$$\min\{X_0^C(k_l)\} > K\max|Q(k,k_l)|, \quad \text{or}$$
$$\frac{M}{N} - (K-1)\frac{M}{N}\mu > K\mu\frac{M}{N}.$$

The condition $1-(K-1)\mu > K\mu$ is equivalent to the well known spark-based condition for the signal reconstruction [58]

$$K < \frac{1}{2}\left(1 + \frac{1}{\mu}\right).$$

From the derivation we can see that this is an extremely pessimistic reconstruction condition.

If we are in position to make a sampling positions strategy, then it should be done in such a way to minimize the value of μ. The minimal value is defined by the Welch bound $\mu \geq \sqrt{\frac{(N-M)}{M(N-1)}}$, [59]. Equality holds for a quite specific form of transforms and measurement matrices (equiangular tight frames - ETF). The DCT does not satisfy the properties of the ETF. Even if it were an ETF, then for $N = 256$ and $M = 128$ condition $\mu \geq \sqrt{(256-128)/128/255} = 0.0626$ holds, according to the Welch bound. The minimal possible value of μ guarantees the recovery for $K < 8.5$. However, as stated before, this is extremely pessimistic for real cases. With $N = 256$ and $M = 128$ we conclude from our numerical analysis that we are able to reconstruct signals with much higher sparsity values. For example, in the worst case of the signal amplitudes and a pessimistic three-sigma rule, we concluded that a full reconstruction with K up to 28 is possible.

IV. Additive Noise influence

Since the signal components have a mean value $\mu_{X_0^C(k_l)} = A_l\frac{M}{N}$, in the reconstruction process they are amplified for N/M in order to produce the correct signal amplitude A_l. Therefore, if there is a small additive noise with variance σ_ε^2 in the signal, its variance will be amplified for $(N/M)^2$ in an initial DCT coefficient estimate

$$\sigma^2_{X_0^C(k)} = \sum_{n\in\mathbf{M}}\sum_{m\in\mathbf{M}} a_k^2 E\{\varepsilon(n)\varepsilon(m)\cos\left(\frac{\pi(2n+1)}{2N}k\right) \times \cos\left(\frac{\pi(2m+1)}{2N}k\right)\} = \frac{M}{N}\sigma_\varepsilon^2.$$

Thus, the variance in one estimated coefficient is

$$\sigma^2_{X_0^C(k)} = \frac{N}{M}\sigma_\varepsilon^2. \quad (19)$$

The reconstructed noise energy in K components, will be

$$E_{\varepsilon R} = \frac{M}{N}\sigma_\varepsilon^2 K\left(\frac{N}{M}\right)^2 = \frac{K}{M}\sigma_\varepsilon^2 N.$$

The signal to noise ratio is

$$SNR = 10\log\left(\frac{E_s}{E_{\varepsilon R}}\right) = 10\log\left(\frac{E_s}{\frac{K}{M}\sigma_\varepsilon^2 N}\right) = 10\log\left(\frac{E_s}{\frac{K}{M}E_\varepsilon}\right) = SNR_i - 10\log\left(\frac{K}{M}\right),$$

where $SNR_i = 10\log(E_s/E_\varepsilon)$ is the input signal to noise ratio in all signal samples.

This result, obtained through a quite simple derivation, will be compared with the one that can be obtained using the Bayesian compressive sensing approach [12]. The covariance

matrix in the estimated coefficients, according to the Bayesian reconstruction approach, is $\Sigma = (\mathbf{A}_{MN}^T\mathbf{A}_{MN}/\sigma_\varepsilon^2 + \mathbf{D})^{-1}$, where $\mathbf{D}$ is a diagonal matrix of hyperparameters. After the positions of nonzero coefficients are found, using an iterative procedure in the Bayesian approach, coefficients with large hyperparameters are excluded along with the corresponding elements of matrix $\mathbf{D}$ and columns of $\mathbf{A}_{MN}$. For our measurement matrix the variance in the estimated coefficients is equal to the diagonal elements of $\mathbf{A}_{MN}^T\mathbf{A}_{MN}$, since the hyperparameters for the nonzero coefficients are zero. Mean value of the diagonal elements of $\mathbf{A}_{MN}^T\mathbf{A}_{MN}$ is M/N. It does not change by omitting columns of the measurement matrix. Therefore, the diagonal elements of the covariance matrix in the final iteration of the Bayesian based reconstruction are $\sigma_\varepsilon^2 N/M$, producing (19).

The presented result for the additive noise influence will be used (and numerically checked) in the analysis of nonsparse signal reconstruction.

V. Nonsparse Signal Reconstruction

Theorem 2: Assume a nonsparse signal with largest amplitudes A_l, $l = 1, 2, ..., K$. Assume that out of total number of N samples only M samples, $1 \ll M \ll N$, are available. Assume that the signal is reconstructed under the assumption as it were K-sparse. The energy of error in the K reconstructed coefficients $\left\|\mathbf{X}_K^C - \mathbf{X}_T^C\right\|_2^2$ is related to the energy of nonreconstructed components $\left\|\mathbf{X}_{T_z}^C - \mathbf{X}^C\right\|_2^2$ as

$$\left\|\mathbf{X}_K^C - \mathbf{X}_T^C\right\|_2^2 = \frac{K(N-M)}{M(N-1)}\left\|\mathbf{X}_{T_z}^C - \mathbf{X}^C\right\|_2^2, \qquad (20)$$

where $\mathbf{X}_K^C$ is a $K \times 1$ vector with the reconstructed coefficients, $\mathbf{X}_T^C$ is a $K \times 1$ vector with the true coefficient values at the reconstructed coefficient positions, $\mathbf{X}^C$ is a $N \times 1$ vector with all true coefficients, and $\mathbf{X}_{T_z}^C$ is a $N \times 1$ vector with K true coefficients at the reconstructed positions and zeros at the remaining $N - K$ positions.

Proof: A nonreconstructed component in the signal behaves as a Gaussian input noise with variance

$$\sigma_{csN}^2 = A_l^2\frac{M(N-M)}{N^2(N-1)}. \qquad (21)$$

All nonreconstructed components will behave as a noise with variance

$$\sigma_T^2 = \sum_{l=K+1}^{N} A_l^2\frac{M(N-M)}{N^2(N-1)}. \qquad (22)$$

After the reconstruction, the total noise energy from the nonreconstructed components (in K reconstructed components) will be

$$\left\|\mathbf{X}_K^C - \mathbf{X}_T^C\right\|_2^2 = E_{\varepsilon R} = K\frac{N^2}{M^2}\sigma_T^2 = \frac{K(N-M)}{M(N-1)}\sum_{l=K+1}^{N} A_l^2.$$

The noise of nonreconstructed components can easily be related to the energy of the nonreconstructed components

$$\left\|\mathbf{X}_{T_z}^C - \mathbf{X}^C\right\|_2^2 = \sum_{l=K+1}^{N} A_l^2$$

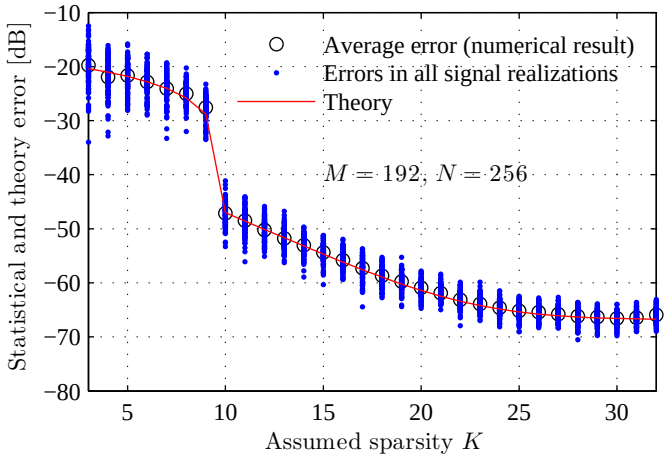

Figure 4. Error energy in the reconstruction of noisy non-sparse signal - calculated numerically and according to the presented theory. Error is shown for various assumed sparsity.

It means that the total error in the reconstructed components is

$$\left\|\mathbf{X}_K^C - \mathbf{X}_T^C\right\|_2^2 = \frac{K(N-M)}{M(N-1)}\left\|\mathbf{X}_{T_z}^C - \mathbf{X}^C\right\|_2^2.$$

This completes the proof.

The previous result can easily be generalized to the noisy signal case. If the input signal contains an input noise whose values are bellow the level of the reconstructed component values in the transformation domain, then

$$\left\|\mathbf{X}_K^C - \mathbf{X}_T^C\right\|_2^2 = \frac{K(N-M)}{M(N-1)}\left\|\mathbf{X}_{T_z}^C - \mathbf{X}^C\right\|_2^2 + \frac{K}{M}\sigma_\varepsilon^2 N. \qquad (23)$$

Example 4: Consider a non-sparse signal:

$$x(n) = \sum_{l=1}^{N} a_{k_l} A_l \cos\left(\frac{\pi(2n+1)}{2N}k_l\right) + \varepsilon(n),$$

with $A_l = 1$ for $l \leq S$ and $A_l = 0.5e^{-2l/(S+1)}$ for $S+1 \leq l \leq N$, all at random DCT indexes $0 \leq k_l < N$. Only $M = 192$ out of $N = 256$ signal samples are available at random positions. Corresponding DCT normalization constants are denoted by a_{k_l}. This signal is approximately S-sparse, with $S = 10$. It is embedded in additive, white, zero-mean Gaussian noise with standard deviation $\sigma_\epsilon = 0.11/N$. Signal was reconstructed using the Algorithm 1 summarized in Appendix B, with various assumed sparsity $3 \leq K \leq 32$. Based on 200 realizations of the signal with random DCT indexes, positions of available samples and random noise realizations, the MSE is calculated and compared with the theoretical result. The error (23) is calculated assuming the normalization to the assumed sparsity. The errors are calculated as follows $E_{numerical} = 10\log\left(\frac{1}{K}\|\mathbf{X}_K^C - \mathbf{X}_T^C\|_2^2\right)$ and $E_{theory} = 10\log\left(\frac{N-M}{M(N-1)}\left\|\mathbf{X}_{T_z}^C - \mathbf{X}^C\right\|_2^2 + \frac{1}{M}\sigma_\varepsilon^2 N\right)$.

The results are presented in Fig. 4. The line represents the theoretical MSE, whereas dots represent the numerical data, whose averaging produces the values indicated by black circles, highly matching the theoretical result.

VI. Application to Audio Signals

The DCT is well known for its applicability in the processing of various signal types: radar, biomedical (ECG, EEG etc.), audio signals, and digital images [14], [23]–[25]. For the context of this paper, especially interesting are the DCT-based algorithms developed for the compression of ECG signals and digital images [14], and for audio signal processing (compression, speech enhancement, denoising, inpainting) [7], [24]–[26], [60], indicating the potential for audio signal representation in the DCT transformation domain with a reduced number of nonzero coefficients. This fact indicates the applicability of sparse signal reconstruction algorithms when the signal samples are missing due to compressive sensing or unavailability, as already confirmed in several recent works [7], [26].

Audio signals are nonstationary with changing spectral content in time. In general they are not sparse [24]. The sparsity can be improved considering localized segments of audio signals [7], [24], [25]. These kind of signals can be then considered as approximately sparse. In order to improve the sparsity of audio signals, a windowed form of the DCT is used, as in the case of MDCT [61]. This form is widely employed in compression procedures involved in modern audio formats [23]. Long duration audio signals $x(n)$ are analyzed with the DCT applied on consecutive blocks of windowed signals

$$x_i(n) = w(n)x(n + iN/2),$$

where $w(n)$ is a windowing function within $0 \leq n \leq N-1$. The subsequent blocks are overlapped such that the second half of one block coincides with the first half of the next block. It is important to note that such a block-based approach in the analysis and processing of audio signals is also important for the presented sparse signal reconstruction algorithms (since it reduces the dimensionality of the partial DCT matrices pseudo-inversion). Note that the block approach is used in the DCT based image analysis as well. If the windowing function form satisfies the condition $w(n) + w(n + N/2) = 1$ within the overlapping interval, $N/2 \leq n \leq N-1$, then the reconstruction of the whole signal is quite simple from the reconstructed windowed signal segments $\hat{x}_i(n)$ as

$$\hat{x}(n) = \sum_i \hat{x}_i(n - iN/2).$$

Many windowing function forms satisfy this condition [14]. The most commonly used windowing function among them is the Hann windowing function $w(n) = 0.5(1 + \cos(\frac{2\pi}{N}(n + \frac{N}{2}))) = \sin^2(\frac{\pi}{N}n)$. In our example the windowing function is used on the analysis side only.

Next, we will assume that a reduced set of distrurbance-free signal samples is available, within a block,

at positions $n \in \mathbf{M} = \{n_1, n_2, ..., n_M\}$. Various circumstances may cause the unavailability of audio signal samples. One illustrative example includes clicks and pops present in the old recordings [24]–[27], [41] highly corrupting certain percent of samples at $n \in \mathbf{Q}$. The set $\mathbf{Q}$ can be considered as a time-domain support function of the localized disturbance. After impulsive disturbances removal, these randomly positioned samples at $n \in \mathbf{Q}$ can are considered as unavailable. They are reconstructed using the presented CS-based method. This issue is illustrated in Examples 5 and 6. Within compressive sensing framework, a reduced set of randomly positioned samples can be initially acquired. This kind of signal is illustrated in Example 7. Nevertheless, both cases, with highly corrupted and omitted randomly positioned signal samples at $n \in \mathbf{Q}$ or with randomly sensed signal samples at $n \in \mathbf{M}$, can be processed in the same way in the reconstruction based on compressive sensing approaches.

Example 5: Embedded MATLAB test signal 'mtlb.mat' is considered, being a low-quality audio recording of a female voice saying the word "matlab", with sampling frequency 7418Hz. Its form is shown in Fig. 5(a). Signal is corrupted with impulsive noise in 15% of randomly positioned samples, Fig. 5(b). Positions of the noise impulses can be easily detected using a limiter (a method for detection of the more complex impulsive noise having amplitudes within the signal values range can be found in [62]). The signal samples at the positions of strong noise are considered as unavailable, and the reconstruction of signal is performed using the rest of samples on blocks, with a Hann windowing function of the length $N = 500$, with overlapping on the half of windowing function length. Reconstruction is preformed using the presented algorithm with various assumed sparsities K. The estimated error in the signal is calculated along with the one presented in Theorem 2. The estimated error is presented by '*' and the one expected by theory is presented by a solid line in Fig. 5(c). The agreement is high. The reconstructed signal segments are added up and the final reconstructed signal is presented in Fig 5(d) for the case of assumed sparsity $K = 150$. RMSE between signals presented in Fig. 5 (a) and Fig. 5(d) is 0.0738.

Example 6: A recorded signal representing word 'Hallelujah' is considered in this example. Signal is recorded on a MacBook computer using MATLAB with sampling frequency 11025Hz. Again we assumed that 20% of arbitrary signal samples are corrupted by a high impulsive noise. These samples are omitted and the signal is reconstructed using the remaining samples only. The result of the reconstruction procedure as in the previous example is presented in Fig 6, where a zoomed signal is also shown for visual clarity of the results.

Example 7: MATLAB signal 'train.mat' is considered in this example, shown in Fig. 7(a). It is an audio recording of a train whistle, sampled at 8192Hz. It has been assumed that the signal is sensed in a compressive way and only 50% of randomly positioned samples are available (as presented in zoomed images in Fig. 7((b) and (c)). The signal is reconstruct assuming various sparsities and the total reconstruction error, is presented in Fig. 7(e). The reconstructed signal with $K = 50$ is shown in Fig. 7(d).

VII. Experimental Evaluation on Audio Signals

Three datasets from [26] are used in the experimental evaluation of the presented theory. Each dataset consists of 10 signals of length 5s from the 2008 Signal Separation Evaluation Campaign [63], [64]. The datasets from this database are:

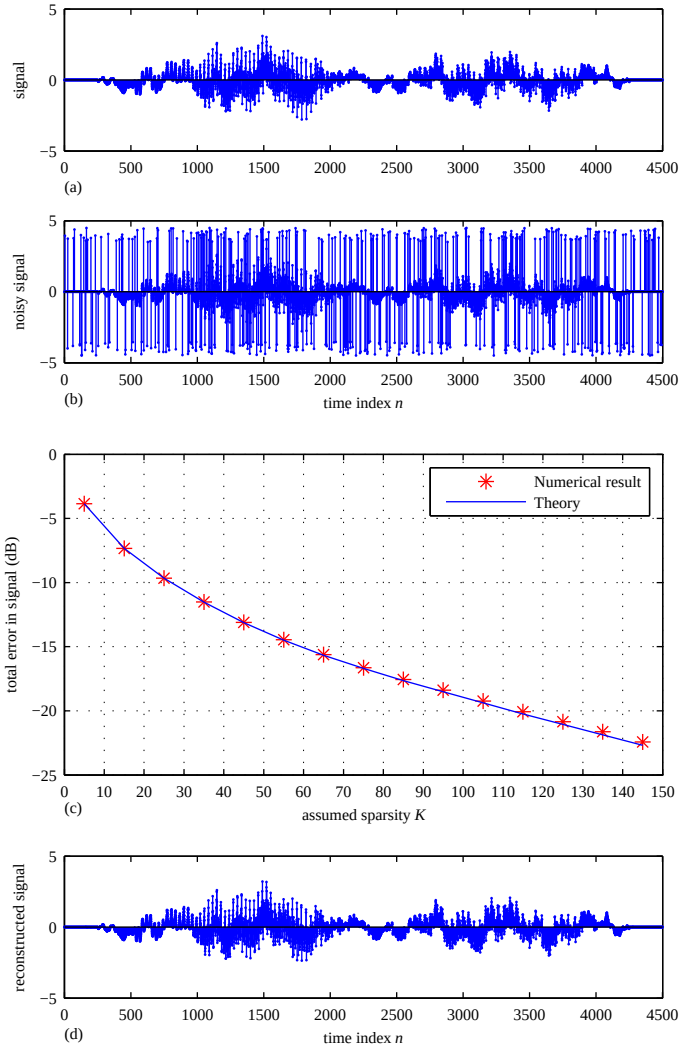

Figure 5. Reconstruction of audio signal 'mtlb.mat' after impulsive noise removal: (a) original signal, (b) signal corrupted with impulsive disturbances, (c) total error energy after the reconstruction with various assumed sparsity, (d) reconstructed signal

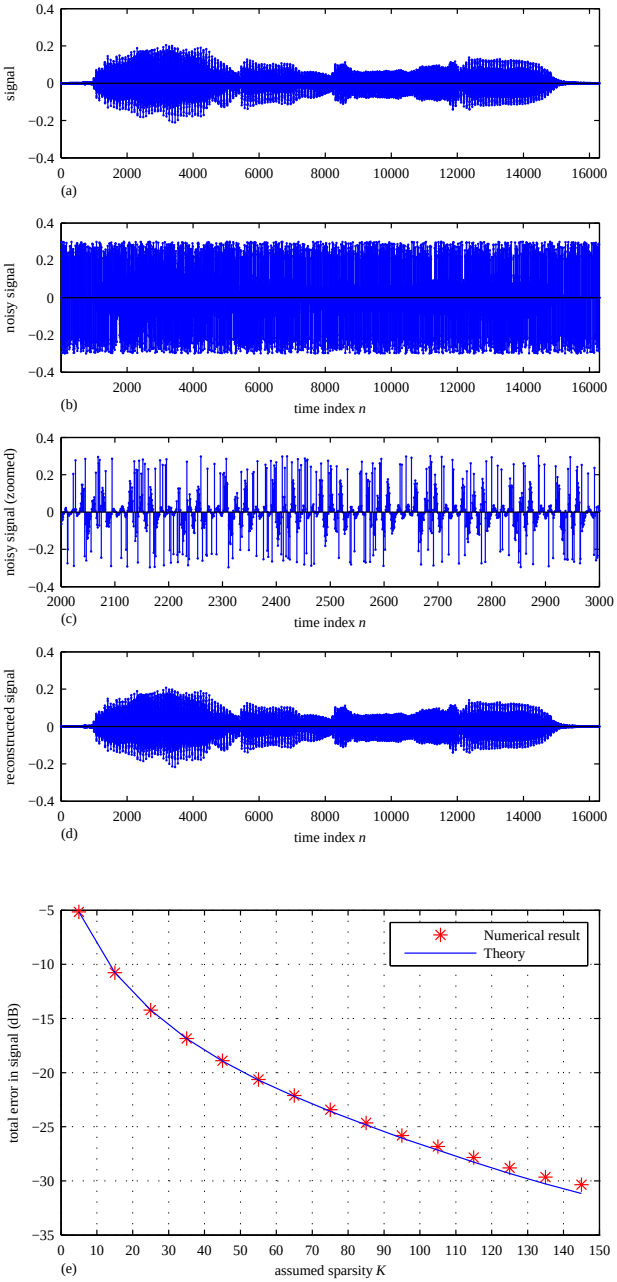

Figure 6. Reconstruction of a recorded audio signal after impulsive noise removal: (a) original signal, (b) signal corrupted with impulsive disturbances, (c) zoomed 1000 samples of the corrupted signal, (d) reconstructed signal, (e) total error energy after the reconstruction with various assumed sparsities.

- Music @ 16kHz: a set of 10 music signals sampled at 16kHz,
- Speech @ 16kHz: consisted of 10 male and female speech signals sampled at 16kHz,
- Speech @ 8kHz: consisted of speech signals sampled at 8kHz, representing a phone quality speech. These signals are obtained by downsampling the signals from the second dataset.

The signals are carefully chosen in order to include a large diversity of audio mixtures and sources. They include both male and female speech from different speakers, singing voice, and pitched and percussive musical instruments [26].

Disturbances are simulated with two possible scenarios. In the first case, it has been assumed that the audio signals are corrupted at random positions. In the second case, the corrupted samples are grouped into random blocks to simulate click/inpainting scenario [41]. In both scenarios the signals are reconstructed and the accuracy of the presented MSE relation is statistically tested.

The presented algorithm is compared with other audio restoration techniques: median and low-pass filtering, then with two widely used model-based audio restoration methods, and with an ℓ_1-norm optimization based reconstruction from the CS framework. The median filters of length 3 and length 5 and low-pass filters of Butterworth type with two cut-off frequences, set in accordance with signals spectrum, are considered. The model-based audio restoration methods are representative examples of the autoregressive (AR) model-based interpolation [27], [36], [66]. Herein we have considered

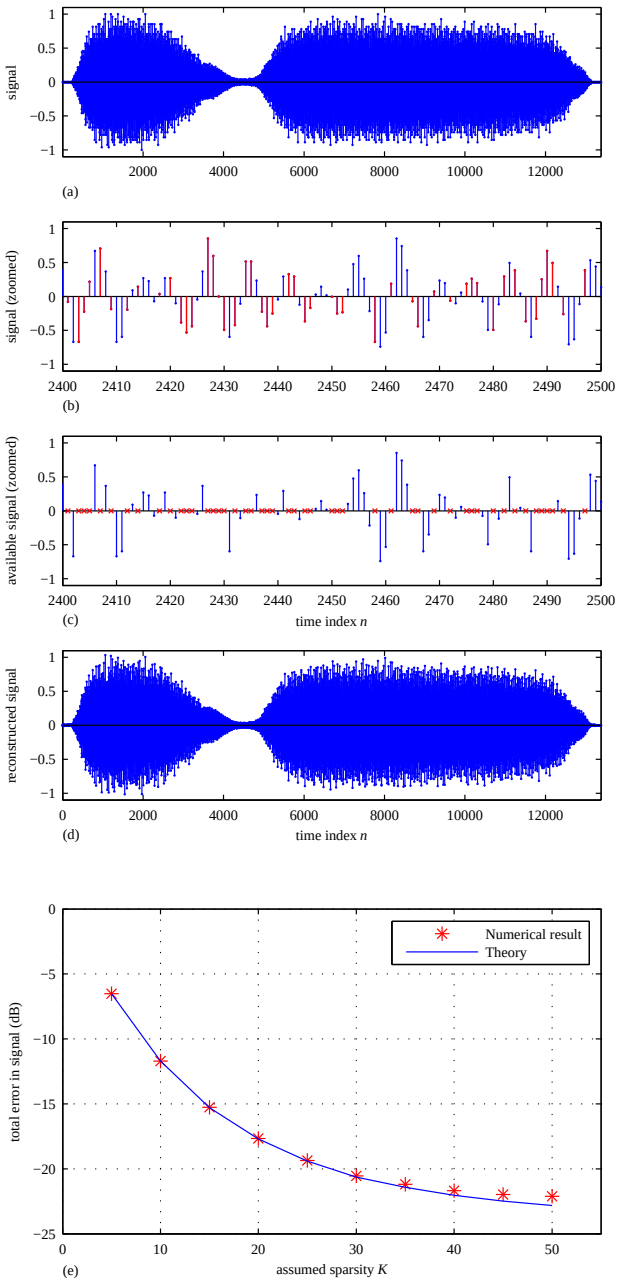

Figure 7. Reconstruction of compressive sensed audio signal having 50% of randomly positioned available samples: (a) original signal, (b) zoomed part of the signal, (c) zoomed part of the signal with crosses at the missing sample positions, (d) reconstructed signal, (e) total error energy after the reconstruction with various assumed sparsities.

the least-squares AR interpolator (LSAR), originally introduced for the concealment of uncorrectable errors in the CD systems [27], [36], along with the implementation provided in the Audio Inpainting Toolbox [26]. The AR model order is set to 30, and the interpolation is performed in blocks of 500 samples, as in the proposed approach. The second considered algorithm is from the class of AR + basis function representation [27], [66]. In our comparative analysis we have used the implementation with sinusoids as basis functions (LSAR+SIN), [66], which has been recently tested in [37], [39]. This algorithm is used with the default settings. The AR model is of order $P = 31$, the number of basis functions is $Q = 31$, and the block size of 1024 samples is used, in accordance with the experimental results presented in [37] and [66]. In all algorithms the reconstruction performance highly depends on the accuracy of the detection of the disturbed sample positions. In order to provide a fair comparison, we have turned off the impulsive disturbances detection in all algorithms, assuming that the disturbed sample positions are correctly detected. In this way, we have been able to test and compare the reconstruction capabilities of these algorithms.

The LASSO-ISTA (Iterative Shrinkage Thresholding Algorithm for LASSO problem) is used as a representative of the base-line ℓ_1-norm minimization reconstruction algorithms [14], [19]. The regularization constant was set to $\lambda = 0.01$.

In the second considered scenario, in addition to these standard methods, two recent approaches that are highly adapted for the removal and reconstruction of clicks in audio signals are also considered [37], [38]. In these two approaches, the reconstruction was performed along with their inherent clicks detection algorithms. However, for the very large number of impulses/blocks in both considered cases, the algorithms failed to perform successful detections. These algorithms (codes) have been used with default settings.

A. Randomly Positioned Disturbances

In the first experiment, signals from each of the three considered datasets are corrupted with a strong impulsive noise in $p\%$ of randomly positioned samples, as illustrated in Example 5 and Fig. 5. These signals are well concentrated in the DCT domain, analyzed in blocks whose length is $N = 500$ samples, weighted with the Hann windowing function. However, they are only approximately sparse. Positions of strong disturbances could easily be detected by using a limiter. More advanced detection methods are described in [27], [37], [38], [41], [62]. Any of them can be used for the detection of corrupted signal samples. Within the CS framework formulation the detected corrupted samples at positions $n \in \mathbf{Q}$ are considered as unavailable measurements. The reconstruction is performed based on the remaining samples in these blocks, considered as the CS measurements at $n \in \mathbf{M}$. Blocks are overlapped by a half of the windowing function length. Reconstruction is done using the presented algorithm with various assumed sparsities K.

1) Theoretical error: First, accuracy of the proposed MSE expression is evaluated. The case with $p = 30\%$ is considered. For the i-th block, the numerical error is calculated as:

$$E^{(i)}_{numerical} = 10 \log \left(\frac{1}{K} \|\mathbf{X}^C_K - \mathbf{X}^C_T\|^2_2 \right) \qquad (24)$$

whereas the theoretical one is given by

$$E^{(i)}_{theory} = 10 \log \left(\frac{N-M}{M(N-1)} \left\|\mathbf{X}^C_{T_z} - \mathbf{X}^C\right\|^2_2 \right). \qquad (25)$$

The squared errors are averaged over blocks, and compared in Fig. 9 as functions of assumed sparsity K, for each signal from each dataset. Solid lines represent the theoretical MSE values whereas the asterisks indicate the numerical results. The agreement of the results is high, confirming the main result of this paper.

2) Comparison with respect to the MSE: The reconstruction results using the considered CS algorithm with sparsity $K = 80$ are compared with the standard approaches for signal filtering and smoothing, then with two standard model-based audio restoration techniques, as well as with an ℓ_1-norm minimization based CS reconstruction with a least absolute shrinkage and selection operator (LASSO) approach [19]–[21]. In this part of the experiment the number of impulses was set to $p = 40\%$.

The results are presented in Table I. The comparison is done with respect to the MSE and objective perceptual quality metrics. These results will be discussed in more details next.

We start the comparison of the presented CS based reconstruction results with a classical method for the impulsive noise reconstruction based on the median filter. The considered filter lengths are 3 and 5. For all three datasets, the median filter of length 3 reduced the MSE for 17.26dB on average, whereas the median filter of length 5 performed similarly (rows denoted as med3 and med5 in Table I).

Next, the low-pass Butterworth filter, as a representative example of a low-pass filtering based smoothing techniques, is considered. Two cutoff frequencies are used on datasets Music @ 16kHz and Speech @ 16kHz. The cutoff frequencies are determined based on the analysis of the signals spectrum. The first filter (row denoted as LPF1) was designed with the normalized cutoff frequency 0.375 and the second one (row LPF2) with 0.625. For the dataset Speech @ 8kHz normalized cutoff frequencies were 0.5, for the first, and 0.7 for the second low-pass filter. Both filters produced similar results, with an MSE drop of 16.25dB on average (as compared to the MSE for the corrupted signal).

The results produced by the state-of-the-art LSAR and LSAR+SIN algorithms were significantly better as compared to the results using the previous classical filtering methods. The average MSE improvement was 25.42dB for the LSAR technique, and 26.55dB for the LSAR+SIN. It is important to note that in the considered scenario, the LSAR implementation [26] suffered from several breakdowns producing high-valued peaks, most likely when the algorithm was not able to track the AR model due to a large number of corrupted/missing values. This hypothesis is confirmed as those breakdowns did not appear in the tests with smaller number of corrupted/missing samples, when the MSE was almost the same as in the LSAR+SIN.

The presented CS method outperfomed the LSAR+SIN for 4dB on average, and the LSAR for about 5dB on average. It is interesting to observe that the improvement was the smallest for the undersampled dataset Speech@8kHz - only 1.69dB. A reason for this could be in the reduced sparsity in the considered domain, arising as a consequence of the signals undersampling.

The ℓ_1-norm minimization based CS reconstruction (the LASSO-ISTA approach) produces results worse than the ones obtained with the presented CS method. These results are worse for about 0.5dB on average than the LSAR results.

3) Comparison with respect to the objective perceptual quality measures: The results presented in Table I are evaluated from the perceptual quality perspective as well. For an objective assessment and prediction of perceived audio quality of signals from Music @ 16kHz dataset, the PEMO-Q method is engaged [67]. We used a freely available online implementation included in the software companion of [68], [69]. We mapped the output of the algorithm, that is, the perceptual similarity measure (PSM_t), to the corresponding objective difference grade (ODG) scale using the procedure described in [67]. The scale has values in the range of -4 (very annoying impairment) to 0 (imperceptible impairment). The PEMO-Q ODG results are calculated for the corrupted signals and for all reconstructed signals, considered in Section VII-A2, for Music @ 16kHz dataset. The results shown in Fig. 8 (first subplot) indicate that the presented CS algorithm outperforms other methods, on average, having an average score of -1.32, compared to the LSAR+SIN score -1.81, and the LASSO-ISTA score -1.85. However, for the 5th and 7th signal the presented method is slightly overperfomed by the LSAR-SIN, and for the 8th and 10th signals it is very slightly outperformed by the LASSO-ISTA algorithm.

Table I
MSE AVERAGED OVER SIGNALS FROM THREE CONSIDERED DATABASES, FOR RANDOMLY POSITIONED IMPULSIVE DISTURBANCES

	Music@16kHz	Speech@8kHz	Speech@16kHz
Noisy	-12.40dB	-25.85dB	-25.60dB
Med3	-27.42dB	-44.04dB	-44.16dB
Med5	-27.41dB	-43.52dB	-44.20dB
LPF1	-26.92dB	-42.42dB	-43.24dB
LPF2	-26.89dB	-42.55B	-43.37dB
LSAR	-36.22dB	-52.58dB	-51.31dB
LSAR+SIN	-37.48dB	-53.59dB	-52.44dB
LASSO	-40.69dB	-49.80dB	-51.15dB
MP	-41.33dB	-55.28dB	-59.02dB

For the perceptual quality comparison of reconstruction results in the speech signals case, the PESQ [65] is used as a quality measure. It is commonly applied in the evaluation of speech quality in the CS-based speech enhancement in various noisy environments [7], [8] with the DCT of windowed audio signal frames as the sparsity domain [7]. Results for the PESQ-based perceptual evaluation obtained for datasets Speech @ 8kHz and Speech @ 16kHz are shown in Fig. 10. The PESQ score is calculated for corrupted signals and for all reconstructed signals considered in Section VII-A2. For Speech @ 8kHz dataset, the PESQ score for the proposed method was 2.89. It is larger than the average LSAR+SIN score 2.59, the LSAR score 2.45, and the LASSO-ISTA score 2.1.

All these algorithms have significantly improved the perceptual quality in comparison to the corrupted signals. The improvement in reconstruction for Speech @ 16kHz dataset is even more evident. In this case the average score of the presented CS method was 3.37. The other considered methods produced following scores: the LSAR+SIN 2.87, the LSAR 2.44, and the LASSO-ISTA 2.27. The improvement is higher than in the case of Speech @ 8kHz dataset.

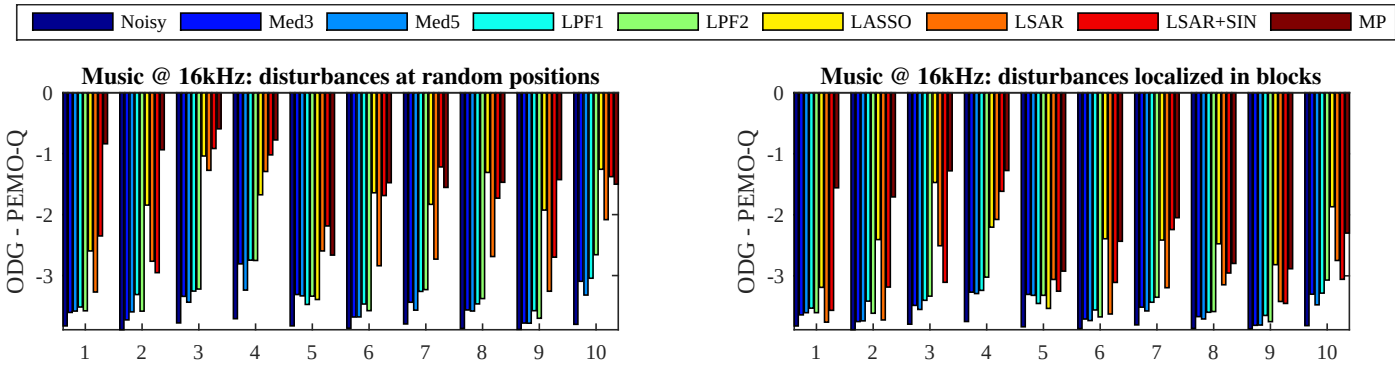

Figure 8. Perceptual evaluation of signals from Music @ 16kHz dataset, using the PEMO-Q ODG metrics. Left figure shows PEMO-Q ODG scores for noisy signal and reconstructed signals in the scenario with randomly positioned impulsive disturbances, corresponding to the MSE results in Table I. Right figure shows results for impulsive disturbances localized in time-domain blocks, with corresponding MSE results shown in Table II.

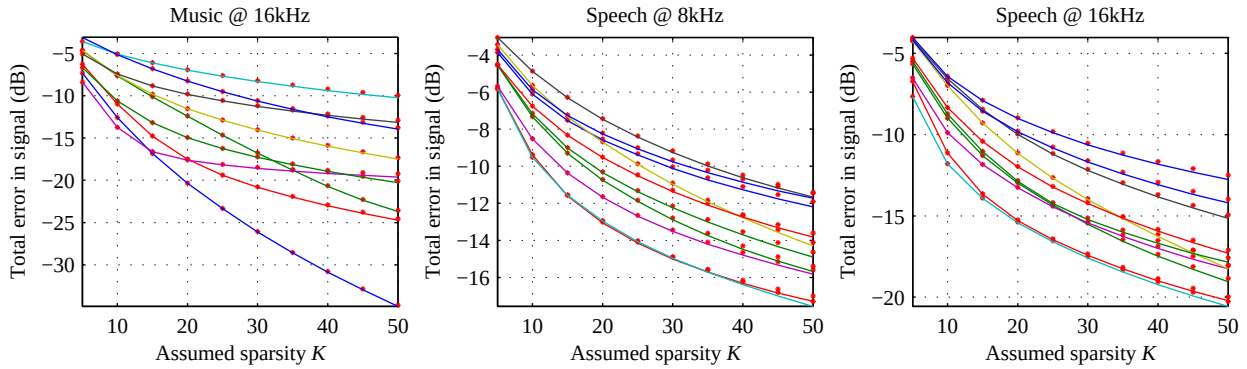

Figure 9. Error energy in the reconstruction of noisy non-sparse signal - calculated numerically (stars) and according to the presented theory (solid lines) for the randomly positioned disturbances. Errors are shown for various assumed sparsities K. First subplot shows the results for 10 music signals sampled at 16kHz, second subplot shows the errors for 10 speech signals sampled at 8kHz whereas the third one shows the errors for 10 different speech signals sampled at 16kHz.

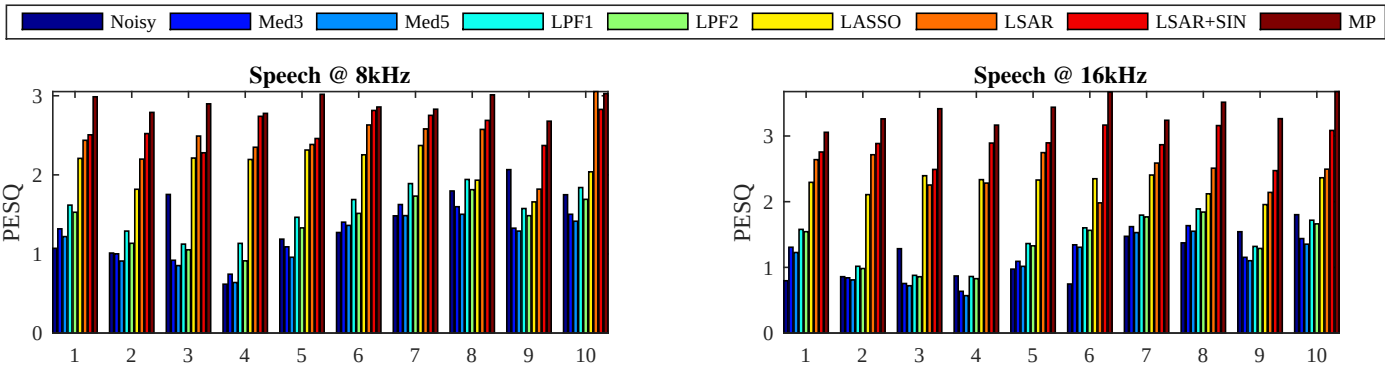

Figure 10. Perceptual evaluation of speech signals using the PESQ metrics, shown for the scenario with randomly positioned disturbances. Signals corrupted with impulsive noise and signals reconstructed using median filters of length 3 and 5, low-pass filters with two cut-off frequencies, LSAR and LSAR+SIN teqniques, LASSO-ISTA CS method and presented MP reconstruction are compared with original non-corrupted signals. The figure is related with the results presented in Table I.

B. Disturbances Localized in the Time-Domain Blocks

In this experiment, impulsive disturbances are located in blocks of subsequent samples. All signals from each of three considered datasets were corrupted with a noise located in randomly positioned blocks having random lengths between 1 and 5, such that in average $p\%$ of samples are affected by the noise. Such impulsive noise is considered in order to simulate the reconstruction potential/performance and theoretical error accuracy in cases when localized time-domain audio signal distortions exist (clicks, CD scratches, clipping etc.), [26], [38], [41]. The reconstruction was performed as in the first experiment, considering half-overlapped signal frames of length $N = 500$, weighted by the Hann windowing function. Corrupted samples are detected and considered as unavailable. They are reconstructed using the presented algorithm with various assumed sparsities.

Table II
MSE AVERAGED OVER SIGNALS FROM THREE CONSIDERED DATABASES FOR THE IMPULSIVE DISTURBANCES APPEARING IN THE TIME-DOMAIN BLOCKS

	Music@16kHz	Speech@8kHz	Speech@16kHz
Noisy	-11.33dB	-23.55dB	-24.53dB
Med3	-26.15dB	-42.70dB	-42.83dB
Med5	-25.96dB	-42.25dB	-42.68dB
LPF1	-25.72dB	-41.62dB	-42.21dB
LPF2	-25.77dB	-41.72dB	-42.31dB
FTR	-26.24dB	-42.92dB	-42.91dB
BDR	-26.66dB	-42.93dB	-42.95dB
LSAR	-32.52dB	-45.48dB	-46.58dB
LSAR+SIN	-34.17dB	-48.61dB	-50.20dB
LASSO	-37.64dB	-48.41dB	-49.65dB
MP	-37.80dB	-50.58dB	-53.17dB

1) Theoretical MSE: Numerically obtained MSE (24) highly matches the theoretical expression (25) in this case as well, as shown in Fig. 11. Results are shown for all signals in all three considered datasets, for the case when the average percent of corrupted samples is $p = 10\%$. Solid lines represent the theoretical error curves, whereas the numerical results are presented using asterisks. The accuracy of the proposed theoretical MSE is expected as long as the conditions for a full CS-based reconstruction are met, even in the cases when the corruption (unavailability) occurs in blocks of successive samples.

2) Comparison with respect to the MSE: The reconstruction results using the presented CS method are compared with the results obtained using the median filters, the low-pass Butterworth filters (with the same parameters as described in Section VII-A2 for the first experimental setup), the LSAR and the LSAR+SIN audio restoration algorithms, and the LASSO-ISTA, with respect to the MSE. The results are presented in Table II, for the case when the average percent of corrupted samples is $p = 50\%$. Additionally, in this scenario the reconstruction MSE is given for a recent method for impulsive noise/clicks detection and removal (AR-based reconstruction) presented in [37], [38]. The detection and the AR-model based reconstruction are done using authors' algorithms, codes and parameters (semi-causal with decision-feedback scheme) [38]. The row denoted by FTR contains the results for the forward-time approach and the row indicated by BDR shows the results with a bidirectional signal processing, originally introduced in [37]. These algorithms are highly adapted for the clicks removal application. For these algorithms the corrupted samples detection was performed using embedded detection procedures. A large number of corrupted samples in this example significantly reduced algorithms' reconstruction efficiency. As in the previous experiment, the considered CS reconstruction techniques, the LSAR, and the LSAR+SIN produced, in average, better results than the other considered methods.

The average MSE improvement with median filters was approximately 17.5dB. It is similar for both filter lengths. The low-pass filtering produced an average improvement of 16.7dB, similar for both considered filters as well. Improvement in the FTR and BDR algorithms was significantly lower for the considered experiment with 50% missing samples than in the case when this percent is lower (e.g. when $p = 10\%$ or 15%). The improvement was on average 17.6dB. The ℓ_1-norm minimization based CS reconstruction (LASSO-ISTA) produced in this experiment better average results, with 25.43dB of MSE improvement, as compared to the LSAR and the LSAR-SIN, producing improvements of 21.72dB and 24.52dB, respectively. The LASSO-ISTA outperformed, on average, the LSAR-SIN due to a significant MSE improvement in the Music @ 16kHz dataset case. The presented CS method produced 1.95dB better result, on average, than the LASSO-ISTA reconstruction. The largest difference in the results occurs for the Speech @ 16kHz dataset.

3) Comparison with respect to the objective perceptual quality measures: For perceptual quality evaluation, the PEMO-Q and PESQ metrics are again used as objective measures. The results for the Music @ 16kHz dataset are shown in Fig. 8 (second subplot). The perceptual evaluation results for the Speech @ 8kHz and Speech @ 16kHz are presented in Fig. 12.

From the PEMO-Q scores in Fig. 8 (second subplot) we may conclude that the presented CS reconstruction, having the average score of -2.12 outperforms the LASSO-ISTA (-2.48), the LSAR+SIN (-2.95), and the LSAR (-3.13). However, for some signals the LASSO-ISTA outperformed the presented CS method. In accordance with the MSE results from Table II, the LASSO-ISTA outperfomed the LSAR and the LSAR+SIN in this experiment.

The PESQ scores, shown in Fig. 12, match the results presented in Table II. For Speech @ 8kHz dataset, the average PESQ score for the presented CS reconstruction is 2.44, outperforming the LSAR+SIN (2.15), the LASSO-ISTA (1.96) and the LSAR (1.83). For Speech @ 16kHz dataset the average scores are: the presented CS reconstruction (2.93), the LSAR+SIN (2.53), the LASSO-ISTA (2.13), and the LSAR (1.68). As indicated in the previous scenario, the perceptual quality improvement is larger for the dataset Speech @ 16kHz.

VIII. CONCLUSION

As one of the most significant signal transforms, incorporated in many compression algorithms, DCT is analyzed here within the framework of a reduced set of observations. As it exhibits many specific properties, the analysis of the DCT is different from the corresponding Fourier analysis. The properties of partial DCT matrix acting as measurement matrix in the considered framework, place it in the middle position between the commonly analyzed partial DFT and Gaussian based measurement matrices. Based on the analysis of the DCT coefficients corresponding to the under-sampled signal, the coherence-based reconstruction condition is derived, with less conservative theoretical bounds guaranteeing successful reconstruction. Additive noise influence on the reconstruction of signals sparse in this particular domain is also analyzed.

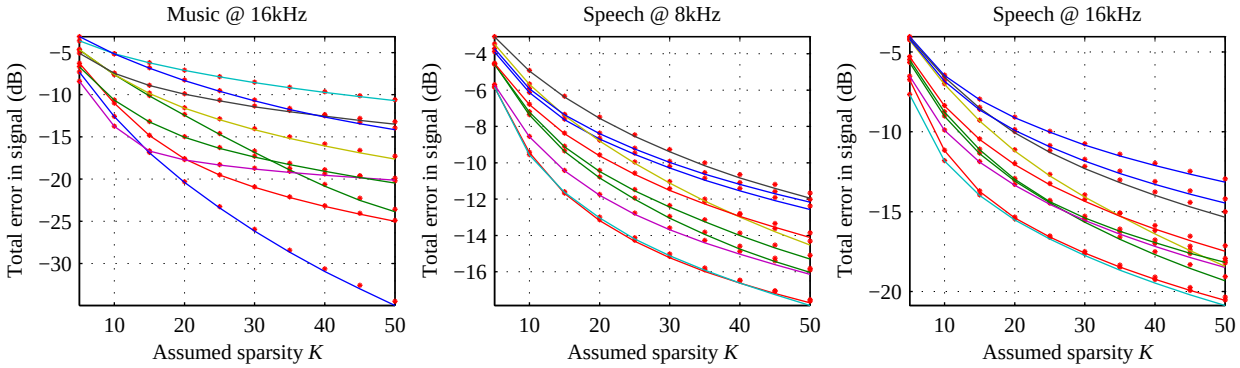

Figure 11. Error energy in the reconstruction of noisy non-sparse signal - calculated numerically (stars) and according to the presented theory (solid lines), when the impulsive noise occurs in the time-domain blocks of varying length. Error is shown for various assumed sparsity.

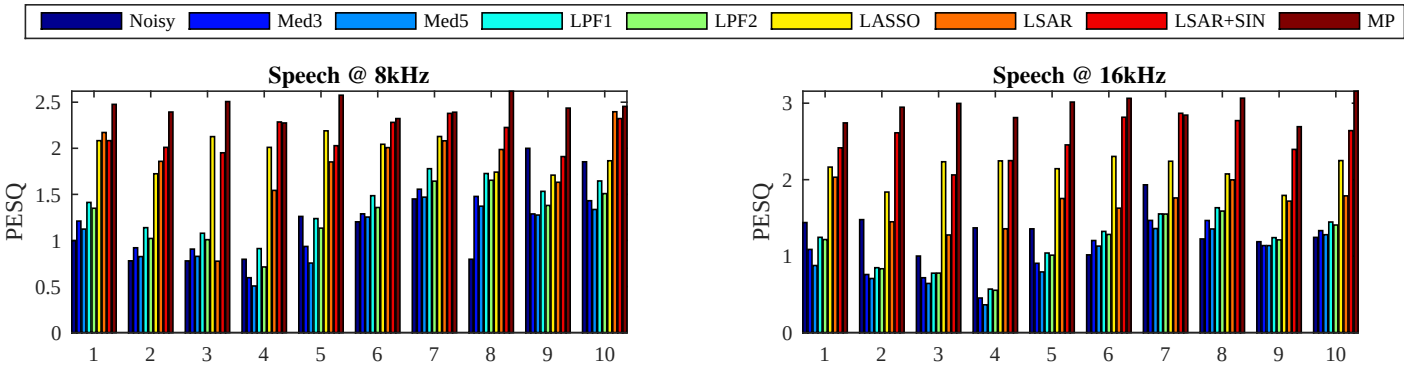

Figure 12. Perceptual evaluation of speech signals using the PESQ metrics shown for the scenario with disturbances localized in the time-domain blocks. Impulsive noise appears in blocks of length varying from 1 to 5 samples. Signals corrupted with impulsive noise and signals reconstructed using median filters of length 3 and 5, low-pass filters with two cut-off frequencies, LSAR and LSAR+SIN teqniques, LASSO-ISTA CS method and presented MP reconstruction are compared with original non-corrupted signals. The figure is related with the results presented in Table II.

Assuming that a nonsparse noisy signal is reconstructed under the sparsity assumption, an explicit analytic expression of the reconstruction error is provided in this paper. A reconstruction algorithm inspired by the presented analysis is proposed. Numerical examples on audio signals confirm the accuracy of the presented theory and efficiency of the reconstruction algorithm as compared to other base-line algorithms for the audio signal reconstructions.

APPENDIX A
PROOF OF THE THEOREM 1

Monocomponent signals: Let us observe a monocomponent signal case, $K = 1$, $k_l = k_1$. Without loss of generality, the amplitude $A_1 = 1$ is assumed. Starting from the theorem assumptions, the initial DCT of a signal with M available samples can be calculated as:

$$X_0^C(k) = \sum_{i=1}^{M} z(k_1, k, n_i). \tag{26}$$

where $z(k_1, k, n_i)$ is defined by (6) with $A_1 = 1$. As the signal and basis functions are orthogonal, it can be written

$$\sum_{n=0}^{N-1} z(k_1, k, n) = \delta(k - k_1). \tag{27}$$

The case for $k = k_1$: For analysis simplicity let us first consider $k = k_1$ and analyze the corresponding DCT coefficient $X_0^C(k_1)$. It is a random variable. According to (27), and due to the fact that all values $z(k_1, k_1, n_i)$ are equally distributed with expected values $1/N$, it can be easily concluded that the mean value of $X_0^C(k_1)$ equals:

$$\mu_{X_0^C(k_1)} = E\left\{X_0^C(k_1)\right\}$$
$$= E\left\{z(k_1, k_1, n_1) + ... + z(k_1, k_1, n_M)\right\} = \frac{M}{N}. \tag{28}$$

For $A_1 \neq 1$ the mean value (28) is multiplied by A_1. Next, we derive the variance of this random variable. It is given by

$$\sigma^2_{X_0^C(k_1)} = E\Big\{ \sum_{i=1}^{M} z^2(k_1, k_1, n_i)$$
$$+ \sum_{i=1}^{M} \sum_{\substack{j=1 \\ i \neq j}}^{M} z(k_1, k_1, n_i) z(k_1, k_1, n_j) \Big\} - \mu^2_{X_0^C(k_1)}. \tag{29}$$

Starting from (27) for $k = k_1$, by multiplying the left and the right side with $z(k_1, k_1, n)$, and taking the expectation of both sides we get:

$$E\left\{z(k_1, k_1, 0) z(n, k_1, k_1) + ... + z(k_1, k_1, N-1) z(k_1, k_1, n)\right\}$$
$$= E\left\{z(k_1, k_1, n)\right\} = \frac{1}{N}. \tag{30}$$

Values $z(k_1,k_1,n)$ are equally distributed. Therefore, expectations $E\{z(k_1,k_1,m)z(k_1,k_1,q)\}$ for $m\neq q,\ m,q\in\mathbf{N}$ are the same and equal to a constant B. We may write:

$$(N-1)\,B+E\left\{z^2(k_1,k_1,m)\right\}=\frac{1}{N} \tag{31}$$

Variance (29) can be now expressed as follows:

$$\sigma^2_{X_0^C(k_1)}=ME\left\{z^2(k_1,k_1,m)\right\}+M\,(M-1)\,B-\frac{M^2}{N^2}, \tag{32}$$

with $m\in\mathbf{M}\subseteq\mathbf{N}$. The expectation appearing in the first term of (32) equals $E\left\{z^2(k_1,k_1,m)\right\}=\frac{a_{k_1}^4}{4}+\frac{a_{k_1}^2}{4N}$ for $k_1\neq 0$. For $k_1=0$ it is equal to $a_{k_1}^4$. Incorporating this result into (32) with B expressed from (31), then multiplying the variance expression with A_1^2 and replacing the values of a_{k_1} we get the result as in Theorem 1:

$$\sigma^2_{X_0^C(k_1)}=\frac{M(N-M)}{N^2(N-1)}\left[1-\frac{1}{2}(1+\delta(k_1))\right]A_1^2. \tag{33}$$

The case $k\neq k_1$: The DCT at non-signal (noisy) positions, $X_0^C(k),k\neq k_1$ is a random variable with statistical properties different from the previously analyzed case. Namely, due to (27) and the fact that all values $z(k_1,k,n_i)$ are equally distributed, it can be concluded that its mean-value is equal to zero, i.e.:

$$\mu_{X_0^C(k)}=E\left\{X_0^C(k)\right\}=0,\ k\neq k_1. \tag{34}$$

For the zero-mean random variable, the variance reads:

$$\sigma^2_{X_0^C(k)}=E\Big\{\sum_{i=1}^{M}z^2(k_1,k,n_i)+\sum_{i=1}^{M}\sum_{\substack{j=1\\ i\neq j}}^{M}z(k_1,k,n_i)z(k_1,k,n_j)\Big\},\ k\neq k_1. \tag{35}$$

Now, starting from (27), multiplying the left and right side by $z(k_1,k,n)$, and taking the expectation of both sides we get similar result as (30) with right side equal to zero. As in the previous case, we may assume that the values of $z(k_1,k,n)$ are equally distributed, and that the expectations $E\left\{z(k_1,k,m)z(k_1,k,q)\right\}$ for $m\neq q,\ m,\ q=0,1,...,N-1$ are the same, and equal to a constant D leading to

$$(N-1)\,D+E\left\{z^2(k_1,k,m)\right\}=0 \tag{36}$$

Since $k\neq k_1$, then the unknown term $E\left\{z^2(k_1,k,m)\right\}$, assuming that $k\neq N-k_1$, can be expressed as

$$E\left\{z^2(k_1,k,m)\right\}=E\left\{z(k_1,k_1,m)\right\}E\left\{z(k,k,m)\right\}=\frac{1}{N^2},$$

$m=0,...,N-1$. According to (27), all values of $z(k_1,k_1,m)$ i.e. $z(k,k,m)$ are equally distributed. It can be easily shown that $E\left\{z(k_1,k_1,m)\right\}=E\left\{z(k,k,m)\right\}=1/N$. For $k=N-k_1$ unknown expectation becomes previously calculated

$$E\left\{z^2(k_1,N-k_1,m)\right\}=E\left\{z^2(k_1,k_1,m)\right\}=\frac{a_{k_1}^4}{4}+\frac{a_{k_1}^2}{4N},$$

$m=0,...,N-1$, for $k_1\neq 0$ and

$$E\left\{z^2(k_1,N-k_1,m)\right\}=a_{k_1}^4,$$

for $k_1=0$. It can be concluded that for the coefficient at the position $k=N-k_1$, the variance expression (33) holds.

Starting from (35), that is $\sigma^2_{X_0^C(k)}=ME\left\{z^2(k,k_1,n_i)\right\}+M\,(M-1)\,D,\ k\neq k_1$ and following the previous conclusions and incorporating the non-zero amplitude $A_1\neq 1$ we get

$$\sigma^2_{X_0^C(k)}=\frac{M(N-M)}{N^2(N-1)}A_1^2\left[1-\frac{1}{2}\delta(k-(N-k_1))\right], \tag{37}$$

where $k\neq k_1$, leading to the result of Theorem 1.

Gaussian distribution: Consider the distribution of $X_0^C(k_1)=\sum_{i=1}^{M}z(k_1,k,n_i)$ for large M and $k_1\neq k$. The probability density function of a normalized zero-mean random variable $c=X_0^C(k_1)/\sigma_{X_0^C(k_1)}$, according to the Edgeworth expression [70], is

$$f(c)=\phi(c)+\frac{1}{4!M}\left[\frac{\kappa_4}{\sigma^4}\phi^{(4)}(c)+\frac{\kappa_3^2}{3\sigma^6}\phi^{(6)}(c)\right]+O(\frac{1}{M^2})(c).$$

The first term is the Gaussian distribution $\phi(c)=e^{-c^2/2}/\sqrt{2\pi}$, while the remaining terms are the deviations from this distribution. The variance, third, and fourth order moments of $z(k_1,k,n_i)$ are denoted by σ^2, κ_3, and κ_4, respectively. In our case, for a large M, we have $\sigma^2\to 1/N^2$, $\kappa_3\to 0$, and $\kappa_4\to 9/(4N^4)$. Therefore $\kappa_4/(4!M\sigma^4)\to 3/(32M)\to 0$ and $f(c)\to\phi(c)$.

Multicomponent signals: The analysis provided for monocomponent signals is extended to the case of multicomponent signals next. The analyzed random variable (26) is now:

$$X_0^C(k)=\sum_{i=1}^{M}\sum_{l=1}^{K}a_k^2A_l\cos\left(\frac{\pi(2n_i+1)}{2N}k_l\right)\times\cos\left(\frac{\pi(2n_i+1)}{2N}k\right). \tag{38}$$

According to the previously presented results, for the case of multi-component signals the DCT coefficients at the l-th signal position $X_0^C(k),\ k=k_l$ behave as random Gaussian variables with non-zero mean values equal to $\mu_{X_0^C(k)}=A_l\frac{M}{N},\ l=1,2,...,K$ whereas the DCT coefficients at non-signal positions, $X_0^C(k),\ k\neq k_l$ also behave as Gaussian variables, with mean-values equal to zero, since the noise caused by missing samples is zero-mean. These conclusions follow from the classical central limit theorem [71], and from the fact that the summation of Gaussian variables produces a new Gaussian variable. The DCT coefficients mean-value for a multicomponent signal can be written as:

$$\mu_{X_0^C(k)}=\frac{M}{N}\sum_{l=1}^{K}A_l\delta(k-k_l).$$

The variance of the DCT coefficients $X_0^C(k)$ at nonsignal positions $k\neq k_l$ equals:

$$\sigma^2_{X_0^C(k)}=\frac{M(N-M)}{N^2(N-1)}\sum_{l=1}^{K}A_l^2\left[1-\frac{1}{2}\delta(k-(N-k_l))\right]. \tag{39}$$

This expression is easily obtained, as at the positions $k\neq k_l$ the missing samples in every signal component contribute to the noise, and the noises from each component

are Gaussian, uncorrelated and zero-mean, with variances $A_l^2 \frac{M(N-M)}{N^2(N-1)} \left[1 - \frac{1}{2}\delta(k-(N-k_l))\right]$, $l = 1, ..., K$, for the noise that originates from the l-th signal component. Note that the result (39) holds in the sense of an average variance of the DCT coefficients at nonsignal positions, as the statistical independence of the random variables is assumed. However, strictly speaking, components of signal multiplied with basis functions may cause a coupling effect if they are placed at positions satisfying certain conditions. For example, in a two-component sparse signal with DCT coefficients at positions k_1 and k_2 if the condition $k_1 + k_2 = 2k$ is satisfied, the coupling effect causes the increase of variances at positions $k_{c1} = (k_1 + k_2)/2$ and $k_{c2} = (k_2 - k_1)/2$. However, at positions $N - k_{c1}$ and $N - k_{c2}$ the variance is decreased for the same values. Consequently, the average variance of DCT coefficients at nonsignal positions $k \neq k_l$ remains the same and equal to (39).

According to the presented analysis for the mono-component signal case, the K-th signal component at the position $k = k_p$, $p \in \{1, 2, ..., K\}$ has variance $A_p^2 \frac{M(N-M)}{N^2(N-1)} \left[1 - \frac{1}{2}(1 + \delta(k_p))\right]$ and mean value $\mu_{X_0^C(k_p)} = A_p^2 M/N$. Additionally, at the position $k = k_p$ the noise caused by missing samples in the remaining $K - 1$ components is also present. This means that the sum of random variables originating from other signal components at positions $k_l, l = \{1, 2, ..., K\}$, $l \neq p$ is added at the position k_p. These $K - 1$ random variables are Gaussian, zero mean, mutually uncorrelated, with variances $A_l^2 \frac{M(N-M)}{N^2(N-1)} \left[1 - \frac{1}{2}\delta(k-(N-k_l))\right]$ and $l \neq p$, with $l = 1, ..., K$, $p = 1, ..., K$. The resulting random variable is also Gaussian, with the mean-value $\mu_{X_0^C(k_p)} = A_p^2 M/N$ and the variance equal to:

$$\sigma^2_{X_0^C(k_p)} = \frac{M(N-M)}{N^2(N-1)} \Big\{ A_p^2 \left[1 - \frac{1}{2}(1 + \delta(k_p))\right] + \sum_{\substack{l=1 \\ l \neq p}}^{K} A_l^2 \left[1 - \frac{1}{2}\delta(k-(N-k_l))\right] \Big\}. \quad (40)$$

Unification of the results given by (39) and (40) leads to Theorem 1 statement for the variance.

APPENDIX B
RECONSTRUCTION ALGORITHM

Inputs to the algorithms are: the available signal samples vector $\mathbf{y}$ and the measurement matrix $\mathbf{A}_{MN}$. The output is the reconstructed signal $\mathbf{x}$. The full inverse transformation matrix $(\mathbf{C}_N)^{-1}$ is used.

1. Iterative algorithm:

$\mathbf{K} = \emptyset, \quad \mathbf{y}_r = \mathbf{y}$
for $i = 1 : K$
$\mathbf{X}_0^C = (\mathbf{A}_{MN}^T \mathbf{A}_{MN})^{-1} \mathbf{A}_{MN}^T \mathbf{y}_r$
$k = \arg\{\max_k |X_0^C(k)|\}$
$\mathbf{K} = \{\mathbf{K}, k\}$
$\mathbf{A}_{MK} = \mathbf{A}_{MN}(K, :)$
$\mathbf{X}_K^C = (\mathbf{A}_{MK}^T \mathbf{A}_{MK})^{-1} \mathbf{A}_{MK}^T \mathbf{y}$
$X_{Kz}^C(k) = X_K^C(k), k \in \mathbf{K}, \quad \text{and} \quad X_{Kz}^C(k) = 0, k \notin \mathbf{K},$
$\mathbf{x}_r = (\mathbf{C}_N)^{-1} \mathbf{X}_{Kz}^C$
$\mathbf{y}_r = \mathbf{y} - \mathbf{x}_r, \quad \text{for} \quad n \in \mathbf{M}$
end
$\mathbf{x} = \mathbf{x}_r$

2. One step algorithm:

$\mathbf{X}_0^C = (\mathbf{A}_{MN}^T \mathbf{A}_{MN})^{-1} \mathbf{A}_{MN}^T \mathbf{y}$
$\mathbf{K} = \arg_k \{(|X_0^C(k)| - 4\sqrt{\frac{(M(N-M)}{N^2(N-1)} \frac{\|\mathbf{y}\|_2^2}{M}}) > 0\}$
$\mathbf{A}_{MK} = \mathbf{A}_{MN}(K, :)$
$\mathbf{X}_K^C = (\mathbf{A}_{MK}^T \mathbf{A}_{MK})^{-1} \mathbf{A}_{MK}^T \mathbf{y}$
$X_{Kz}^C(k) = X_K^C(k), k \in \mathbf{K}, \quad \text{and} \quad X_{Kz}^C(k) = 0, k \notin \mathbf{K},$
$\mathbf{x} = (\mathbf{C}_N)^{-1} \mathbf{X}_{Kz}^C$

The algorithms can be combined.

ACKNOWLEDGMENT

This work is supported by the Montenegrin Ministry of Science, project grant: CS-ICT New ICT Compressive sensing based trends applied to: multimedia, biomedicine and communications (Grant No. 01-1002).

REFERENCES

[1] D. Donoho: "Compressed sensing," *IEEE Trans. on Information Theory*, 2006, vol. 52, no. 4, pp. 1289–1306
[2] R. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.
[3] E. J. Candès, M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*,Vol. 25, Issue 2, pp. 21–30, 2008.
[4] L. Stanković, S. Stanković, and M. Amin, "Missing Samples Analysis in Signals for Applications to L-estimation and Compressive Sensing,"*Signal Processing*, vol. 94, pp. 401-408, Jan 2014.
[5] M. Elad, *Sparse and Redudant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.
[6] H. Rauhut, "Stability Results for Random Sampling of Sparse Trigonometric Polynomials," *IEEE Trans. on Information theory*, 54(12), pp. 5661–5670, 2008
[7] D. Wu, W. P. Zhu and M. N. S. Swamy, "The Theory of Compressive Sensing Matching Pursuit Considering Time-domain Noise with Application to Speech Enhancement," *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 22, no. 3, pp. 682-696, March 2014.
[8] J. C. Wang, Y. S. Lee, C. H. Lin, S. F. Wang, C. H. Shih and C. H. Wu, "Compressive Sensing-Based Speech Enhancement," *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 24, no. 11, pp. 2122-2131, Nov. 2016.
[9] R. E. Carrillo, K. E. Barner, and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE Journal of Selected Topics in Signal Processing* , 2010, vol. 4, no. 2, pp. 392–408.

[10] E. Sejdić, M. A. Rothfuss, M. L. Gimbel, M. H. Mickle, "Comparative Analysis of Compressive Sensing Approaches for Recovery of Missing Samples in an Implantable Wireless Doppler Device," *IET Signal Processing*, vol. 8, no. 3, pp. 230-238, May 2014.

[11] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: algorithms and applications," *IEEEAccess*, vol. 3, pp. 490–530, 2015.

[12] S. Ji, Y. Xue, L. Carin, "Bayesian Compressive Sensing", *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[13] S. Stanković, I. Orović, and L. Stanković, "An Automated Signal Reconstruction Method based on Analysis of Compressive Sensed Signals in Noisy Environment," *Signal Processing*, vol. 104, Nov 2014, pp. 43 – 50, 2014.

[14] L. Stanković, *Digital Signal Processing with Selected Topics*, CreateSpace Independent Publishing Platform, An Amazon.com Company, November 4, 2015

[15] L. Stanković, I. Stanković, and M. Daković, "Nonsparsity Influence on the ISAR Recovery from Reduced Data," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 52, Issue: 6, Dec. 2016, pp. 3065 - 3070

[16] G. Bi, S. K. Mitra, S. Li, "Sampling rate conversion based on DFT and DCT," *Signal Processing*, Volume 93, Issue 2, pp. 476–486, Feb. 2013.

[17] L. Zhao, G. Bi, L. Wang, H. Zhang, "An Improved Auto-Calibration Algorithm Based on Sparse Bayesian Learning Framework," *IEEE Signal Processing Letters*, vol. 20, no.9, pp. 889–892, 2013.

[18] G. Davis, S. Mallat and M. Avellaneda, "Greedy adaptive approximation,"*Journal of Constructive Approximation* , vol. 12, pp. 57–98, 1997.

[19] R.J. Tibshirani, "Regression shrinkage and selection via the lasso: A retrospective," *Journal of the Royal Statistical Society* , Series B, 73:273–282, 2011.

[20] A.Beckand, M.Teboulle,"A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, no. 2, pp. 183202. 2009.

[21] L. Stanković, and M. Daković, "On a Gradient-Based Algorithm for Sparse Signal Reconstruction in the Signal/Measurements Domain," *Mathematical Problems in Engineering*, vol. 2016, Article ID 6212674, 11 pages, 2016. doi:10.1155/2016/6212674.

[22] M. G. Christensen, J. Østergaard and S. H. Jensen, "On compressed sensing and its application to speech and audio signals" *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, 2009, pp. 356–360.

[23] V. Britanak and K. R. Rao, "An efficient implementation of the forward and inverse MDCT in MPEG audio coding," *IEEE Signal Processing Letters*, vol. 8, no. 2, pp. 48–51, Feb. 2001.

[24] P. Maechler et al., "VLSI Design of Approximate Message Passing for Signal Restoration and Compressive Sensing," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 2, no. 3, pp. 579–590, Sept. 2012.

[25] D. Bellasi, P. Maechler, A. Burg, N. Felber, H. Kaeslin and C. Studer, "Live demonstration: Real-time audio restoration using sparse signal recovery," *2013 IEEE International Symposium on Circuits and Systems (ISCAS2013)*, Beijing, 2013, pp. 659-659.

[26] A. Adler, V. Emiya, M. G. Jafari, M. Elad, R. Gribonval and M. D. Plumbley, "Audio Inpainting," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 3, pp. 922-932, March 2012. doi: 10.1109/TASL.2011.2168211

[27] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration–A Statistical Model-Based Approach,* Berlin, Germany: Springer-Verlag, 1998.

[28] W. Etter, "Restoration of a discrete-time signal segment by interpolation based on the left-sided and right-sided autoregressive parameters," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1124-1135, May 1996.

[29] Han Lin and S. Godsill, "The multi-channel AR model for real-time audio restoration," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2005., New Paltz, NY, 2005, pp. 335-338.

[30] P. J. W. Rayner and S. J. Godsill, "The Detection and Correction of Artefacts in Degraded Gramophone Recordings," *Final Program and Paper Summaries 1991 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, 1991, pp. 151-152.

[31] S. J. Godsill and P. J. W. Rayner, "A Bayesian approach to the restoration of degraded audio signals,"*IEEE Transactions on Speech and Audio Processing*, vol. 3, no. 4, pp. 267-278, Jul 1995.

[32] S. J. Godsill and C. H. Tan, "Removal of low frequency transient noise from old recordings using model-based signal separation techniques," *Proceedings of 1997 Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, 1997, pp. 4

[33] S. J. Godsill and P. J. W. Rayner, "Robust noise modelling with application to audio restoration," *Proceedings of 1995 Workshop on Applications of Signal Processing to Audio and Accoustics*, New Paltz, NY, 1995, pp. 143–146.

[34] C. M. Hicks and S. J. Godsill, "A two-channel approach to the removal of impulsive noise from archived recordings," Acoustics, Speech, and Signal Processing", *1994. ICASSP-94., 1994 IEEE International Conference on, Adelaide*, SA, 1994, pp. II/213-II/216 vol.2.

[35] S. J. Godsill, P. J.Wolfe, and W. N. W. Fong, "Statistical model-based approaches to audio restoration and analysis," *J. New Music Res.*, vol. 30, no. 4, pp. 323–328, 2001.

[36] A. Janssen, R. Veldhuis and L. Vries, "Adaptive interpolation of discrete-time signals that can be modeled as autoregressive processes" *in IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 317-330, Apr 1986.

[37] M. Niedźwiecki and M. Ciołek, "Elimination of Impulsive Disturbances From Archive Audio Signals Using Bidirectional Processing," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 5, pp. 1046-1059, May 2013.

[38] M. Ciołek and M. Niedźwiecki, "Detection of impulsive disturbances in archive audio signals," *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, 2017, pp. 671-675.

[39] M. Ruhland, J. Bitzer, M. Brandt and S. Goetze, "Reduction of Gaussian, Supergaussian, and Impulsive Noise by Interpolation of the Binary Mask Residual," *in IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1680–1691, Oct. 2015.

[40] M. Siu and A. Chan, "A Robust Viterbi Algorithm Against Impulsive Noise With Application to Speech Recognition," *in IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 2122-2133, Nov. 2006. doi: 10.1109/TASL.2006.872592

[41] F. R. Avila and L. W. P. Biscainho, "Bayesian Restoration of Audio Signals Degraded by Impulsive Noise Modeled as Individual Pulses," *in IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 9, pp. 2470-2481, Nov. 2012. doi: 10.1109/TASL.2012.2203811

[42] J. K. Nielsen, M. G. Christensen, A. T. Cemgil, S. J. Godsill and S. H. Jensen, "Bayesian interpolation in a dynamic sinusoidal model with application to packet-loss concealment," *2010 18th European Signal Processing Conference*, Aalborg, 2010, pp. 239-243

[43] H. Ofir, D. Malah and I. Cohen, "Audio Packet Loss Concealment in a Combined MDCT-MDST Domain," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 1032–1035, Dec. 2007.

[44] H. Ofir and D. Malah, "Packet Loss Concealment for Audio Streaming based on the GAPES and MAPES Algorithms," *2006 IEEE 24th Convention of Electrical & Electronics Engineers in Israel*, Eilat, Israel, 2006, pp. 280–284.

[45] C. Perkins, O. Hodson and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *in IEEE Network*, vol. 12, no. 5, pp. 40-48, Sept.-Oct. 1998.

[46] D. Goodman, G. Lockhart, O. Wasem and Wai-Choong Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 6, pp. 1440-1448, Dec 1986.

[47] B. K. Lee and J. H. Chang, "Packet Loss Concealment Based on Deep Neural Networks for Digital Speech Transmission," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 2, pp. 378-387, Feb. 2016.

[48] A. Stenger, K. Ben Younes, R. Reng and B. Girod, "A new error concealment technique for audio transmission with packet loss," *1996 8th European Signal Processing Conference (EUSIPCO 1996)*, Trieste, Italy, 1996, pp. 1-4.

[49] B. W. Wah, Xiao Su and Dong Lin, "A survey of error-concealment schemes for real-time audio and video transmissions over the Internet," *Proceedings International Symposium on Multimedia Software Engineering*, Taipei, 2000, pp. 17-24.

[50] C. A. Rodbro, M. G. Christensen, S. V. Andersen and S. H. Jensen, "Compressed domain packet loss concealment of sinusoidally coded speech" *Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03). 2003 IEEE International Conference on*, 2003, pp. I-104–7 vol.1.

[51] Subramanya et al., "Automatic Removal of Typed Keystrokes," *IEEE Signal Proc. Letters*, Vol. 14, No. 5, May 2007

[52] J. Gemmeke, H. Van Hamme, B. Cranen, and L. Boves, "Compressive sensing for missing data imputation in noise robust speech recognition," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 272–287, Aug. 2010.

[53] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," *IEEE Transactions on Speech and Audio Processing*, 2(2):345-349, 1994.
[54] O. Cappé and J. Laroche, "Evaluation of short-time spectral attenuation techniques for the restoration of musical recordings," *IEEE Transactions on Speech and Audio Processing*, 3(1):84-93, 1995.
[55] S. Canazza, G. De Poli and G. A. Mian, "Restoration of Audio Documents by Means of Extended Kalman Filter, " *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 6, pp. 1107-1115, Aug. 2010. doi: 10.1109/TASL.2009.2030005
[56] H. Buchner, J. Skoglund and S. Godsill, "An acoustic keystroke transient canceler for speech communication terminals using a semi-blind adaptive filter model," *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, 2016, pp. 614-618.
[57] V. Britanak, P. C. Yip and K. R. Rao, *Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations*, Academic Press& Elsevier Science, Amsterdam, 2007.
[58] D. Donoho and M. Elad., "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Natl. Acad. Sci.*, 100(5):2197–2202, 2003.
[59] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inform. Theory*, 20(3), pp. 397–399, 1974
[60] J. Huang and Y. Zhao, "A dct-based fast signal subspace technique for robust speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 8, no. 6, pp. 747–751, Nov. 2000.
[61] H. S. Malvar, *Signal Processing with Lapped Transforms*, Artech House, Boston, 1992.
[62] L. Stanković, M. Daković, and S. Vujović, "Reconstruction of Sparse Signals in Impulsive Disturbance Environments," *Circuits, Systems and Signal Processing*, vol. 2016. pp. 1-28, DOI: 10.1007/s00034-016-0334-3, ISSN: 0278-081X print, 1531-5878 online
[63] E. Vincent, S. Araki, and P. Bofill, *The 2008 Signal Separation Evaluation Campaign: A Community-Based Approach to Large-Scale Evaluation*. Paraty, Brazil: Springer, Mar. 2009.
[64] Audio signals database [available online]: http://small-project.eu/software-data, last accessed Sept. 2017
[65] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra, "Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs," *in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 2, 2001, pp. 749–752.
[66] J. Nuzman, "Audio restoration: An investigation of digital methods for click removal and hiss reduction," *University of Maryland, Institute for Advanced Computer Studies*, 2004 [Online]. Available: www.github.com/jnuzman/audio-restoration-2004, last updated/accessed in Dec. 2017.
[67] R. Huber and B. Kollmeier, "PEMO-Q–A New Method for Objective Audio Quality Assessment Using a Model of Auditory Perception," *in IEEE Transactions on Audio, Speech, and Language Processing*, vol. 14, no. 6, pp. 1902–1911, Nov. 2006.
[68] V. Emiya, E. Vincent, N. Harlander and V. Hohmann, "Subjective and objective quality assessment of audio source separation," *IEEE Transactions on Audio, Speech and Language Processing*, 19(7):2046-2057, 2011.
[69] E. Vincent, "Improved perceptual metrics for the evaluation of audio source separation, " *10th Int. Conf. on Latent Variable Analysis and Signal Separation (LVA/ICA 2012)*, 2012.
[70] M. Hazewinkel (Ed.) *Edgeworth series*, Encyclopedia of Mathematics, Springer, 2001.
[71] P. Billingsley, *Probability and Measure*, Third ed., John Wiley & sons, 1995.

Ljubiša Stanković (*M'91–SM'96–F'12*) was born in Montenegro in 1960. He received the B.S. degree in EE from the University of Montenegro (UoM), the M.S. degree in Communications from the University of Belgrade and the Ph.D. in Theory of Electromagnetic Waves from the UoM. As a Fulbright grantee, he spent 1984-1985 academic year at the Worcester Polytechnic Institute, USA. Since 1982, he has been on the faculty at the UoM, where he has been a full professor since 1995. In 1997-1999, he was on leave at the Ruhr University Bochum, Germany, supported by the AvH Foundation. At the beginning of 2001, he was at the Technische Universiteit Eindhoven, The Netherlands, as a visiting professor. He was vice-president of Montenegro 1989-90. During the period of 2003-2008, he was Rector of the UoM. He was Ambassador of Montenegro to the UK, Ireland, and Iceland from 2010 to 2015. His current interests are in Signal Processing. He published about 450 technical papers, more than 150 of them in the leading journals, mainly the IEEE editions. Prof. Stanković received the highest state award of Montenegro in 1997, for scientific achievements. He was an Associate Editor of the *IEEE Transactions on Image Processing*, the *IEEE Signal Processing Letters*, *IEEE Transactions on Signal Processing*, and numerous special issues of journals. Prof. Stanković is a Senior Area Editor of the *IEEE Transactions on Image Procesing*, Associate Editor of the *IET Signal Processing*, and a member of Editorial Board of *Signal Processing*. He is a member of the National Academy of Science and Arts of Montenegro (CANU) since 1996 and a member of the European Academy of Sciences and Arts. Stanković (with coauthors) won the Best paper award from the European Association for Signal Processing (*EURASIP*) in 2017 for a paper published in the Signal Processing journal.

Miloš Brajović (*S'12*) was born in Podgorica, Montenegro, in 1988. He received the B.S. and M.Sc. degrees in electrical engineering from the University of Montenegro, Podgorica, Montenegro, in 2011 and 2013, respectively. He is currently working toward the Ph.D. degree at the University of Montenegro. He is currently working as a teaching assistant at the University of Montenegro. He is a member of the Time-Frequency Signal Analysis Group, University of Montenegro, where he is involved in several research projects. His research interests include signal processing, time-frequency signal analysis, and compressive sensing. He has published several papers in these areas.