CS Performance Analysis for the Musical Signals Reconstruction

Marijana Šćekić, Radomir Mihajlović, Irena Orović, Srdjan Stanković
Faculty of Electrical Engineering
University of Montenegro
Podgorica, Montenegro
e-mail: irenao@ac.me

Abstract— **The Compressive Sensing (CS) method for reconstruction of musical signals is analyzed in this paper. CS is a new method for signal acquisition which has been developed in recent years. In the CS scenarios, it is possible to reconstruct the entire signal information from just a small set of randomly chosen measurements, using different minimization algorithms. Consequently, this method founds application in a large number of signal processing areas. The analyzed musical signals and the applied acquisition procedure, satisfy two important CS requirements. Namely, the observed signals have sparse representation in frequency domain, and the measurement procedure provides conservation of the main information about the signal, despite the reduction of the number of analyzed samples. Musical signals of different nature and complexity are observed in the paper. The efficiency of the CS reconstruction is analyzed for different number of available measurements. It will be shown that the minimal number of measurements required for successful signal reconstruction depends on the complexity of musical tones. Based on reconstruction error, the simple CS procedure for classification of two types of musical signals is presented. The reconstruction accuracy is measured by mean relative error between original and reconstructed signal, as well as perceptually – by listening both original and reconstructed signal.**

Keywords - Compressive Sensing, signal classification, signal reconstruction, string instruments, wind instruments

I. INTRODUCTION

Nowadays, the musical signals storage requires significant memory capacities, especially in the cases of significant number of samples that have to be recorded. A large number of compression algorithms exist for musical signals and they are still intensively developed. However, another challenging task that arisen in recent years is the possibility to directly acquire musical signals in compressive manner. In other words, it is interesting to analyze the possibility and the efficiency of applying Compressive Sensing (CS) signal reconstruction to the randomly acquired small set of musical samples.

CS [1]-[4] is introduced in order to represent a signal using small number of linear measurements, i.e., signal samples. The number of samples is usually much smaller than it is required when the signal is sampled at the Nyquist rate. It is important to emphasize that signal must fulfil certain requirements in order to be analyzed by using the CS. First important CS requirement refers to a measurement procedure. Namely, the signal has to be measured in a way that minimizes the number of samples necessary for signal reconstruction and analysis, i.e., measurements have to be incoherent. Furthermore, the signal has to be sparse in the certain domain (Fourier, discrete cosine transform, wavelet, etc.). For sparse signals, the important information about the signal is condensed into the small number of non-zero coefficients. The signal measurements in the CS procedure are acquired from the domain where the signal has dense representation. An example of sparse signal is sinusoid – although sinusoidal signal has non-zero samples most of time, in frequency domain it is represented by peak at a certain frequency. Therefore, it can be said that sinusoid has sparse representation in the frequency domain. When talking about musical signals, we can say that they consist of or can be modeled using small number of time-varying sinusoidal signals [5], [6]. For that reason, musical signals exhibit sparsity property in the frequency domain, and consequently, they can be good candidates for the applications of CS approach. The characteristics of musical signals, including their complexity, depend on the tone or sound quality, the frequency of the sound wave, the musical instrument on which sound is played, and so on. Different nature of musical signals causes different number of harmonics, which directly affects their sparsity. Accordingly, the musical signal complexity will affect the required number of Compressive Sensed signal samples that can still provide successful reconstruction of the entire signal information. Based on the minimal number of samples used for signal reconstruction, in this paper we define a simple procedure for the classification of some musical tones. The musical tones reconstruction is tested for different number of randomly acquired measurements. The reconstruction accuracy is quantified using the mean relative error between the original and reconstructed signal. The quality of reconstructed signal is also proved by using the listening tests.

This work has been supported by the project CS-ICT (New ICT Compressive sensing based trends applied to: multimedia, biomedicine and communications), funded by Montenegrin Ministry of Science.

The paper is organized as follows: In Section II basic concepts of CS method are given. The procedure for CS musical signal reconstruction is explained more detailed in Section III. The experimental results and error analysis, as well as error-based classification are given in Section IV. Conclusion is given in Section V.

II. THEORETICAL BACKGROUND

Sampling the signals according to the sampling theorem results in large amount of information that has to be further processed. However, the number of important features within this large number of data are, in most cases, much smaller. Therefore we might say that these signals are compressible. This feature of real signals is being exploited by the CS [7], [8]. The idea is to reduce the number of samples during acquisition, so that there is no need for further signal compression. However, complex and sophisticated mathematical algorithms are needed for the entire signal reconstruction and analysis.

Let us first explain term "sparse signal". The *N*-length signal is said to be sparse in certain basis if it can be represented by *K* non-zero samples in the frequency domain, where *K*<<*N* holds. If signal *f* is finite-length signal, it can be represented as [9]-[11]:

$$f = \psi x \, , \tag{1}$$

where ψ is an $N \times N$ matrix whose columns are the basis functions ψ_i, x is the ψ – transformed vector and it is sparse in domain ψ. Clearly, f and x are equivalent representations of the one-dimensional signal. By using properly chosen measurement matrix $\theta_{M \times N}$, the vector of measurements could be written as:

$$b = \theta \psi x \, , \tag{2}$$

where vector $b_{M \times 1}$ contains CS measurements. Relation (2) could be written in the form:

$$b = Ax \, , \tag{3}$$

where $A_{M \times N}$ is a CS matrix. The CS case of interest is when *M*<<*N* holds, i.e. vector of acquired samples used in measurement procedure should be as small as possible.

The equation (3) has no unique solution. In order to reconstruct f with good accuracy, eq. (3) must meet two conditions: it has to be sufficiently sparse and the matrices must be incoherent. Random matrices are largely incoherent with any fixed basis, and thus are used as measurement matrices in CS. The system could be solved by using the optimization algorithms [12], [13], such as greedy algorithms (MP, OMP, StOMP, CoSaMP, etc.), convex relaxation algorithms and the least absolute shrinkage and selection operator (LASSO).

III. COMPRESSIVE SENSING BASED RECONSTRUCTION OF MUSICAL SIGNALS

Sound generation in acoustic musical instruments is based on the vibrations of strings (e.g. violin, cello, piano), or air, membranes (e.g. flute, clarinet). These vibrations appear on certain frequencies. Vibrations of strings (air or membranes) on one frequency lead to vibrations on the whole number of frequencies that are multiplies of that frequency, called fundamental frequency or pitch. These multiplies of the fundamental frequency are called harmonics (overtones). Having in mind harmonic nature of the musical signals, they can be good candidates for CS reconstruction.

The attention is paid on the number of harmonics, i.e., number of sinusoidal components in the musical signals produced by various instruments. Signal is reconstructed with CS method, by using certain percent of the total signal length. In order to provide sparse signal representation, we have used discrete cosine transform (DCT) as transform basis. Depending on the signal complexity, reconstruction is done with 25% or 35% of the total signal length. Measurements are taken randomly from the dense domain, i.e., time domain of the signal. For all considered signals, non-noisy environment is assumed. The CS measurement matrix *A* is, in fact, submatrix of the matrix ψ. It is obtained by using set of rows from the DCT matrix ψ. Which row to take is obtained with random permutations of vector *d*, described as follows:

$$d = P(N), \tag{4}$$

where *P* denotes random permutations of *N* elements and *N* is the signal length. Matrix A is then obtained as:

$$A = \psi(1:d\left(1:M\right),1:N)) \tag{5}$$

As it can be seen from (5) only *M* rows are used, where *M* is defined number of measurements. The undetermined system of equations,

$$b = \theta \psi x = \psi(1:d\left(1:M\right),1:N))x \, , \tag{6}$$

is solved by using l_1 – minimization [13]-[16], described as:

$$\min_x \|x\|_{l_1} \quad subject \ to \ b = Ax \, . \tag{7}$$

In this paper, the reconstruction is done by l1-minimization, using basis pursuit (BP) algorithm. The complexity of the BP depends on the type of dictionary we are using and has range from *O(n)* to *O(n log (n))*, where *n* is signal length [16].

IV. ANALYSIS AND EXPERIMENTAL RESULTS

In this section the results of musical signals reconstruction are given. Several types of real musical signals are observed: five signals belong to the string instruments, and another five signals belong to the wind instruments group. As previously mentioned, the different instruments may sound different even if they have the same fundamental frequency. This difference is caused by a difference in overtones - frequencies present above the fundamental frequency that sometimes are not at exactly integer multiples of the fundamental. As a consequence, the sound has a unique characteristics from instrument to instrument. In order to demonstrate different nature of the musical instruments, in all considered cases we have observed one note. In the cases of the musical signals created by several instruments, signal sparsity is degraded and CS could fail to reconstruct signal exactly.

In all considered examples total signal length is chosen to be 6000 samples, and sparse signal domain is DCT domain (for matrix ψ). As representatives of the two groups, the violin (string instruments) and the trumpet (wind instrument) are used. The signal reconstruction is tested for different number of measurements. As violin has more overtones compared with the trumpet, it will require 35% of samples for successful reconstruction, while this number is 25% for the trumpet signal. The number of measurements is chosen experimentally. The chosen number of measurements is minimal number which gives no audible signal degradation after reconstruction. This is proved by listening both, the original and reconstructed signals. Original and reconstructed DCTs of the violin signal are shown in Fig. 1, while Fig. 2 shows zoomed original (blue) and reconstructed (red) signal in the time domain. Frequency and time domains of the original and reconstructed trumpet signals are show in Figs. 3 and 4, respectively.

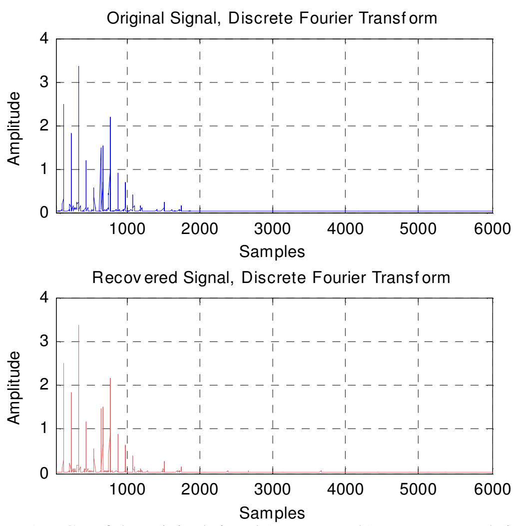

Figure 1. DCT of the original signal (upper graph); reconstructed signal (lower graph) for violin signal

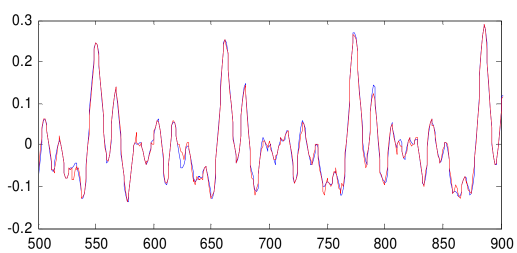

Figure 2. Original (blue) and reconstructed (red) violin signal, time domain

Comparing the frequency domains of the observed signals it is clear that trumpet signal is more sparse, i.e. it has less non-zero frequency components, for the same played note. Therefore, smaller number of measurements is required for trumpet than for violin signal reconstruction, in order to obtain reconstructed signal of the same quality.

Reconstruction accuracy is also proved numerically by calculating mean relative errors between original and reconstructed signal. The errors are calculated in cases when 25% of samples are used in reconstruction and the results are given in the Table 1.

To achieve better visual comparison between the observed signals, time-frequency (TF) representations of the signals are calculated. As TF representation, the S-method is used. It is defined as [17]:

$$SM(t,\omega)=\frac{1}{2\pi}\int_{-\infty}^{\infty}P(\theta)STFT(t,\omega+\frac{\theta}{2})STFT^{*}(t,\omega-\frac{\theta}{2})d\theta, \quad (8)$$

where $P(\theta)$ is the frequency window function and $STFT$ denotes Short-Time Fourier transform. S-methods of the observed signals are presented in Fig. 5.

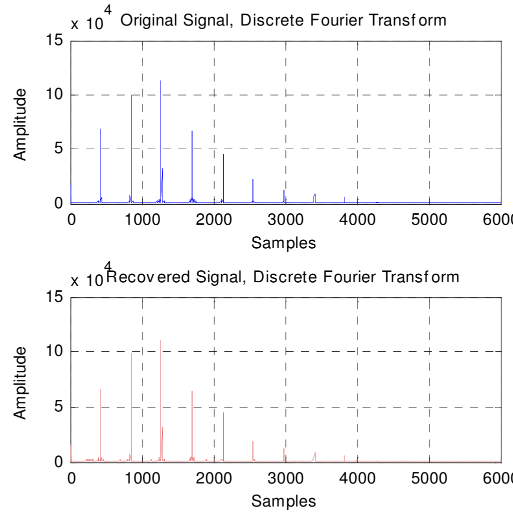

Figure 3. DCT of the original signal (upper graph); reconstructed signal (lower graph) for trumpet signal

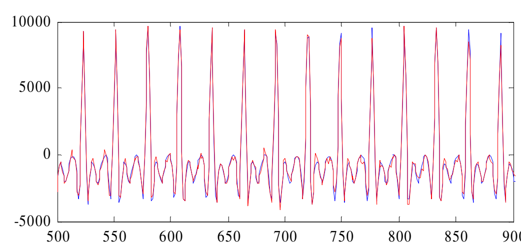

Figure 4. Original (blue) and reconstructed (red) trumpet signal, time domain

From Fig. 5 it can be seen that violin signal has large number of closely spaced components in the TF plane, while the distance between harmonics in the trumpet signal is larger. This proves that the trumpet signal is more sparse, as we noticed by observing DCT domains of the signals.

A. *CS based classification of the musical signals*

Different number of samples is required for high quality signal reconstruction in different musical signals groups. In this part, we have observed reconstruction accuracy using the same number of measurement for all considered signals. It is shown that the classification of the signals, based on reconstruction error, could be performed.

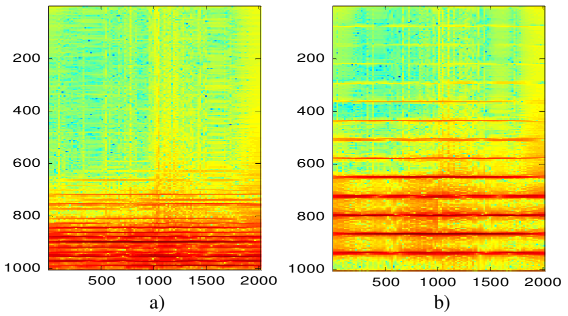

Figure 5. S-method of the a) violin; b) trumpet signal

Two groups of the signals are observed: string and wind instrument groups. The reconstruction with the 25% of samples is performed. The reconstruction with this number of samples gives no audible distortion for wind instruments, while for the string instruments gives small degradations, but still high quality reconstruction. In the Table 1 the mean relative errors of the reconstruction are given. The errors are shown to be of 10^{-3} order for wind instruments group, while for the string instruments this error is larger - 10^{-2} or 10^{-1} order. Note that these reconstruction errors do not produce audible signal degradation. Therefore, signals could be classified into string and wind group, based on the mean relative reconstruction error. Let us note that the experiments for each signal are repeated several times. The errors in the Table 1 are the average of the errors obtained in each repetition.

TABLE I. MEAN RELATIVE ERRORS BETWEEN ORIGINAL AND RECONSTRUCTED SIGNALS

Type	Instrument	Error
String instruments	VIOLA	0.02280
	VIOLIN	0.01000
	CELLO	0.05800
	PIANO	0.83000
	GUITAR	0.02430
Wind instruments	FLUTE	0.00734
	TRUMPET	0.00175
	CLARINET	0.00389
	OBOE	0.00493
	TROMBONE	0.00855

V. CONCLUSION

CS method application on the musical signals of different types and complexity is analyzed in this paper. Audio signals with different number of sinusoidal components are observed. The 25% and 35% of samples is used as number of measurements, depending on the signal complexity. This number of taken measurements is shown to give satisfactory results, i.e. reconstruction with no audible distortion. The reconstruction accuracy is proved by listening signals, as well as by calculation mean relative errors between original and reconstructed signal. It is shown that signals which consist of smaller number of overtones could be reconstructed by using smaller number of measurements (25% for the wind instruments group). Also, the classification of the signals between two groups – wind and string instruments, could be performed, based on the reconstruction error, when the same number of measurements is used in reconstruction. Namely, measured mean relative errors between original and reconstructed signals are shown to be of 10^{-3} order for wind, and $10^{-2}/10^{-1}$ order for string instruments.

REFERENCES

[1] S. Stankovic, I. Orovic, E. Sejdic, "Multimedia Signals and Systems", ch.6., Springer-Verlag, New York, 2012.
[2] R. Baraniuk, "Compressive Sensing", Lecture Notes in IEEE Signal Processing Magazine vol. 24, July 2007.
[3] J. Haupt, R. Castro, R. Nowak, G. Fudge, A. Yeh "Compressive sampling for signal classification", Fortieth Asilomar Conference on Signals, Systems and Computers, 2006. ACSSC '06. pp. 1430-1434, Oct. 29 2006-Nov. 1 2006.
[4] A. Cohen, W. Dahmen, R. DeVore, "Compressed sensing and best *k*-term approximation", July 24, 2006.
[5] I. Orovic, S. Stankovic, A. Draganic, "Time-Frequency Analysis and Singular Value Decomposition Applied to the Highly Multicomponent Musical Signals," Acta acustica united with acustica , vol. 100 (2014) 1, Accepted for publication, 2013.
[6] A. Griffin, T. Hirvonen, C. Tzagkarakis, A. Mouchtaris, P. Tsakalides, "Single-Channel and Multi-Channel Sinusoidal Audio Coding Using Compressed Sensing," IEEE Transactions on Audio, Speech, and Language Processing, vol. 19, no. 5, pp.1382,1395, July 2011.
[7] S. Stankovic, I.Orovic, M. Amin, "Compressed sensing based robust time-frequency representation for signals in heavy tailed noise", ISSPA 2012, Canada 2012.
[8] L. Stankovic, S. Stankovic, M. Amin, "Missing Samples Analysis in Signals for Applications to L-estimation and Compressive Sensing," Signal Processing, vol. 94, Jan 2014, pp. 401-408, 2013.
[9] M. Elad, "Sparse and Redudant Representations: From Theory to Applications in Signal and Image Processing," Springer 2010.
[10] E. J. Candès, M. B. Wakin, "An Introduction To Compressive Sampling," IEEE Signal Processing Magazine 2008.
[11] D. Donoho, "Compressed sensing," IEEE Transaction on Information Technology, vol. 52, no. 4, 2006, pp. 1289 – 1306.
[12] L. Stankovic, S. Stankovic, I. Orovic, M. Amin, "Robust Time Frequency Analysis based on the L-estimation and Compressive Sensing," IEEE Signal Processing Letters, vol. 20, no. 5, pp. 499-502, 2013.
[13] S. Stankovic, I. Orovic, L. Stankovic, "Single-Iteration Algorithm for Compressive Sensing Reconstruction," 21st Telecommunications Forum TELFOR 2013, Novembar, 2013.
[14] I. Orovic, S. Stankovic, B. Jokanovic, A. Draganic, "On Compressive Sensing in Audio Signals," ETRAN 2012, Zlatibor, 2012.
[15] P. Boyd, L. Vandenberghe, "Convex Optimization", Cambridge University Press, Mar 8, 2004 - Business & Economics - 716 pages.
[16] D. L. Donoho, Y. Tsaig, "Fast Solution of l1-norm Minimization Problems When the Solution May be Sparse", IEEE Transactions on Information Theory, Nov. 2008.
[16] S. S. Chen, D. L. Donoho, M. A. Saunders. "Atomic Decomposition by Basis Pursuit," SIAM Journal on Scientific Computing 20(1):33-61, January 1998.
[17] S. Stankovic, I. Orovic, N. Zaric, "Robust speech watermarking procedure in the time-frequency domain," EURASIP Journal on Advances in Signal Processing, vol. 2008, no. ID 519206, Page(s) 9 pages, 2008.