# Empirical Investigation of Compressed Sensing applicability to Lossy Audio Compression

Rubem J. V. de Medeiros, Edmar C. Gurjão and João M. de Carvalho
Electrical Engineering Department
Federal University of Campina Grande
Campina Grande, PB. Braxil
kukamed@gmail.com, ecandeia@dee.ufcg.edu.br, carvalho@dee.ufcg.edu.br

***Abstract*— Compressive sampling is a new framework that exploits sparsity of a signal in a transform domain to perform sampling below the Nyquist rate. In this paper we investigate the applicability of the Compressed Sensing Framework to audio compression by searching for a good sparsity basis and a reconstruction technique fit to audio applications. We also propose a new method for lossy audio compression of real, non-sparse audio signals, based on our investigations. The method uses the Modified Discrete Cosine Transform (MDCT) as a sparse basis and the l-1 norm optimization for signal reconstruction. We evaluate final audio quality with the Perceptual Evaluation of Audio Quality (PEAQ) algorithm. The method we propose has the properties of reverse-complexity, cryptography, error-resiliency and universality of encoder, altogether without any additional hardware. .**

***Keywords*— *sampling; audio; compressed; quality.***

## I. Introduction

The well know Shannon/Nyquist Sampling Theorem states that in order to perfectly reconstruct a periodic band limited signal, it should be sampled with a rate at least twice its highest frequency [1], [2]. Consequently, all sampling hardware design for any class of signals (audio, video, speech, MRI, RF, etc) obeys the Shannon/Nyquist Theorem in order to guarantee lossless reconstruction.

The core tenet of signal processing is that signals often contain some type of structure that enables efficient representation (compression) and processing. For example, transforms such as the Discrete Fourier Transform (DFT), the Discrete Cosine Transform (DCT), the Short Time Fourier Transform (STFT) and the Discrete Wavelet Transform (DWT), exploit structures of signals of dimension $N$ for sparse representation in a $K$ dimension space. Therefore, if we have the transform of the signal, we can transmit it by sending only the $K \ll N$ transform coefficients.

The classical approach for signal processing systems is to perform sampling, obeying Shannon/Nyquist Theorem [2], [1], and compress the samples afterwards. CandÃ¨s, Romberg and Tao [3] and Donoho [4] proposed a novel approach, Compressed Sensing (CS), by which sampling a signal, sparse or compressible in some basis, is a linear random projection. CS combines steps of sampling and compressing in order to sample below the Nyquist rate, without aliasing or frequency loss. With CS framework are associated the properties of universality, error-resiliency, cryptography and reverse-complexity [4].

In this work, based on perceptual audio quality, we empirically analyze the performance of a system that applies compressed sensing to audio signals. This paper is organized as follows: in Section II we make a review of the main concepts behind Compressed Sensing; in Section III we present related works found in the literature; the proposed Audio Compressed Sensing method is described in section IV; the results obtained with the proposed compression are presented in Section V and in Section VI we present some conclusions and propose future works.

## II. Compressed Sensing

Considering the data of interest a real-valued unknown vector $\mathbf{x} \in \mathbb{R}^N$, we acquire linear measurements $\mathbf{y} = \Phi\mathbf{x}$, where $\Phi$ is the $M \times N$ measurement matrix and $\mathbf{y} \in \mathbb{R}^M$. Since $\Phi$ maps vectors in $\mathbb{R}^N$ to vectors in a smaller dimensional space $\mathbb{R}^M$, in order to recover signal $\mathbf{x}$ from $\mathbf{y}$, additional information is needed. To understand these additional information we will review some concepts.

### A. Sparsity

*Definition 1:* A vector $\mathbf{f} \in \mathbb{R}^N$ is S-sparse on basis $\Psi$ if $||\Psi\mathbf{f}||_0 = S$

The $l_0$ norm of a vector, denoted by $||x||_0$ is the number of non-zero components in this vector. Thus, $||\Psi\mathbf{f}||_0$ is the number of non-zero components in the projection of $\mathbf{f}$ on basis $\Psi$.

*Definition 2:* The base $\Psi$ is a sparsity basis of $\mathbf{f} \in \mathbb{R}^N$ if $\mathbf{f}$ is S-sparse and $S \ll N$.

As signals can be represented on different basis, if we are able to find sparse basis for a certain class of signals, we can limit their sparsity to a maximum value $K$. However, signals of practical interest (image, audio, video) are almost sparse since their components decay rapidly and we can discard small coefficients with small loss on perceptual quality. Once small components are forced to be zero, the new signal is strictly sparse. We call almost sparse signals compressible signals. Audio signals are compressible signals.

### B. Incoherence

*Definition 3:* Let $\Psi, \Phi \in \mathbb{R}^{\sqrt{N} \times \sqrt{N}}$ be orthonormal basis. The coherence between these basis is given by:

$$\mu(\Psi, \Phi) = \sqrt{N} \max\{\Phi^T \Psi\}. \qquad (1)$$

In other words, the coherence between two basis is the highest value of all inner products between elements of $\Psi$ and $\Phi$. If we have small coherence between $\Phi$ and $\Psi$, a signal that is sparse on $\Phi$ domain is not sparse on $\Psi$ and vice-versa. In a non-sparse representation, each element of a vector carries information of any other element. Therefore, a small portion of global signal information lies on every sample. A good measurement basis must be incoherent with any sparsity basis, so Bernoulli and Gaussian matrices are well suited for that task [5].

### C. Reconstruction

CS framework constrains the original signal to be sparse in some sparsity basis. Thus, the reconstruction techniques will use optimization in order to search for a vector that is the most sparse possible. Two basic reconstruction algorithms are used to recover the original signal: Basis Pursuit [3] and Orthogonal Matching Pursuit [6]. Basis pursuit tries to find a vector with lower $l_1$ norm that satisfies the CS measurement. Orthogonal Matching Pursuit tries to find each component individually, subtracting the contribution of each component via the greedy method. Best reconstruction results are achieved by Basis Pursuit, but Orthogonal Matching Pursuit is the fastest between the two algorithms.

## III. Related Works

To the best of our knowledge, only two works with applications of Compressed Sensing (CS) to audio signals are found in the literature. Griffin and Tsakalides [7] have investigated the best sparsity basis and reconstruction algorithms for real audio signals from sensor networks, concluding that DCT is the best one for most signals (except impulsive signals), and Multiple Sensor Basis Pursuit is the best method for reconstruction for audio multiple sensor networks. As an application, they proposed an audio sensor network as a location detection system. They used the Signal-to-Distortion Rate (SDR) as an objective measurement of audio quality, obtaining good SDR values.

Carmi, Kanevsky, and Ramabhadran [8] presented an algorithm for lossy speech compression based on CS-based Kalman Filtering that exploits sparsity of audio signal on the DFT domain. Results are presented only as spectral plots of original and reconstructed speech signals that do not tell much about the quality of the technique.

## IV. Audio Compressed Sensing

In order to adapt the Compressed Sensing framework to audio signals, we followed the steps described in this section.

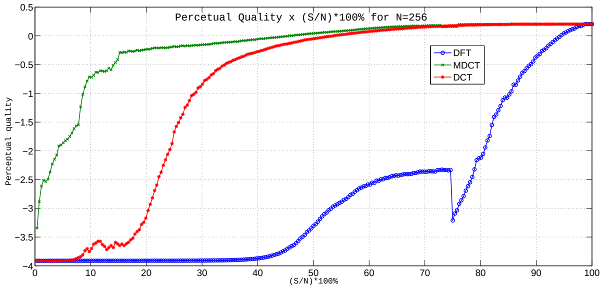

Fig. 1. Perceptual Audio Quality versus Sparsity rate S/N with size of window N=512.

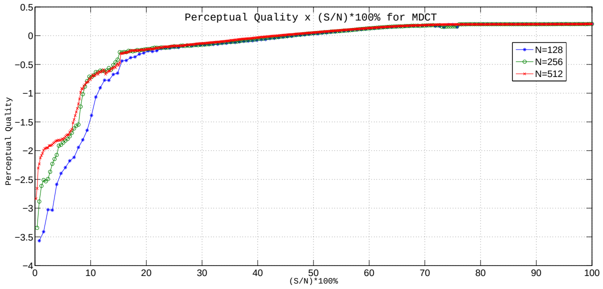

Fig. 2. Perceptual Audio Quality versus Sparsity rate S/N for different window size N of the MDCT.

### A. Investigation of the best audio sparsity basis

We realized a simulation to compare Discrete Fourier Transform, Discrete Cosine Transform and the Discrete Modified Cosine Transform as audio sparsity basis. For each audio window, we truncated the signal sparsity and measured its Objective Difference Grade by the PEAQ [9] algorithm. We tested 10 seconds long audio samples from the sound quality evaluation material of EBU (*European Broadcasting Union*) [15]. The following samples were tested:

- pop music: ABBA (Stereo).wav
- classical music: Piano (Schubert) (Stereo).wav, Clarinet (arpegio;melodious phrase) (Stereo).wav, Orchestra (R. Strauss) (Stereo).wav
- speech: Female Speech (English) (Mono).wav, Male Speech (English) (Mono).wav

The results can be seen in figure 1. From this figure, we can conclude that among the tested basis, the MDCT is the best for audio sparsity basis. We also tested the influence of the window size on the perceptual quality (figure 2), observing that for sparsity levels below 20% large window sizes improve quality.

### B. Investigation of the reconstruction technique

We evaluated the performance of a $l_1$ minimization algorithm. We fixed the frame size (N =128, 256, 512) and varied the sparsity rate $(S/N)$. For each sparsity rate $(S/N)$, we generated 100 random signals and tested for how

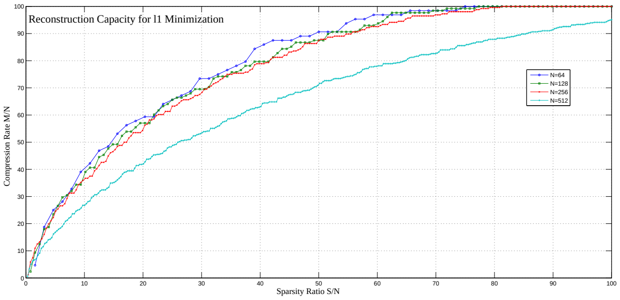

Fig. 3. Reconstruction from $l_1$ norm minimization - Compression rate M/N x Sparsity rate.

many measurements $M/N$ we had a perfect recovery[1] of the original signal. A perfect recovery guarantees that the algorithm performs a perfect recovery with $M$ measurements for input signal of sparsity $S$. The results are shown in figure 3. For example, for a frame size $N = 256$, to perform a perfect reconstruction of a signal with sparsity rate approximately $S/N = 0.18 \Longrightarrow S = 46$, we will have to measure approximately $M/N = 0.5 \Longrightarrow M = 128$ samples.

### C. Description of the Lossy Audio Compression System

*1) Encoder:* For simplicity, our encoder consists of a sampling matrix. The chosen sampling matrix was the pseudo-random Bernoulli matrix. Note that CS measurement needs $M$ measurements to work, and so the analogous operation is equivalent to multiplying the discrete audio signal sampled above the Nyquist rate by a Bernoulli matrix with each line corresponding to a sample. The operation of multiplying the audio signal by a pseudo-random sequence must occur above the Nyquist rate so that the operation is equivalent to the same operation on the discrete domain. Observe that the encoder demands a simple hardware, since it only performs the product of the signal vector by the sensing matrix, which implies in low computational complexity on the encoder side. This is an interesting property for systems that cannot dispose of complex acquisition hardware, typical of Wireless Sensor Networks [14].

The encoder can also be designed for software implementation. In that case, the encoder operation would be just the multiplication between the measurement matrix and the discrete above-Nyquist sampled audio data, also blocked in frames. This operation has a very low computational complexity, characterizing a low-complexity encoder as well.

*2) Decoder:* The decoder consists of basically two functional blocks: the $l_1$ minimization and the inverse transform. The $l_1$ minimization block performs the following operation:

$$\min_{\tilde{\mathbf{x}} \in \mathbb{R}^n} \|\tilde{\mathbf{x}}\|_1 \text{ subject to } \|\mathbf{\Phi}\mathbf{\Psi}\tilde{\mathbf{x}} - \mathbf{y}\|_2 < \varepsilon \tag{2}$$

Where $\mathbf{\Psi}$ is the sparsity basis for audio signals and $\mathbf{\Phi}$ is the measurement basis used by the encoder.

In other words, the decoder finds the $\tilde{\mathbf{x}}$ with lowest $l_1$ norm and, when applied CS, gives the same $\mathbf{y}$ as from actual measurements. This operation can recover the original audio data $\mathbf{x}$ almost exactly (error bounded by $\varepsilon$) or partially, depending on the sparsity level $K$ and the number of measurements.

In our decoder, the $l_1$ minimization finds $\tilde{\mathbf{x}}$ in the sparsity basis domain, that is the transform domain, so we need to apply the inverse transform to obtain the signal on the time domain. Since the transform adopted is the MDCT, as it is a lapped transform, a final step of overlap and add is needed.

Several properties come along with the Compressed Sensing framework that is of direct interest to the data compression community.

*3) Error Resiliency:* Due to noise, errors may happen when Bob transmits audio messages to Alice, so Alice may receive corrupted samples. After decoding, the message is surprisingly not lost, but only received with low quality. This happens because CS measurements present error-resiliency, since random measurements spread global information among all samples, so each sample carries the same amount of information, for a certain level of quality. Therefore we can protect the audio from errors by increasing the number of random measurements. The properties of error resiliency and error correction are detailed in [12] [13]

*4) Universal Encoder:* Although Compressed Sensing is not adaptive, it is universal in the sense that for any class of audio signals the encoder will be exactly the same. In CS, the sparsity level is the important parameter, and not the band or where and how the frequency components are distributed. As CS acquires signal by random matrices that have low coherence with any other basis, no matter what sparsity basis is used in the decoder (different sparsity basis for different class of audio signals), the measurement and sparsity basis will be incoherent. The same encoder can encode any audio data (speech, solo instrument, music, etc)and the design of the decoder will be different for differs classes of signals. The property of universality by random projections is discussed in [11] [5].

*5) Reverse Complexity:* Classically, the encoder is the component with high computational cost, while the decoder is typically low-cost. In our proposed model, the encoder has very low-complexity and the decoder has a much larger complexity. That property is suitable for applications where a low-cost encoder is needed. However, in audio data applications, sensors need a larger amount of memory resources to be able to store the collected data and it is highly ineffective to transmit raw data. Thus the proposed technique can reduce the amount of memory needed and the amount of data to transmit.

## V. Evaluation of Perceptual Audio Quality of the Proposed System

The proposed system using as sparsity base the MDCT, as sampling base the Bernoulli matrix and $l_1$ minimization as reconstruction technique, was tested with the same samples from European Broadcasting Union (EBU). We can

[1] We assume a perfect recovery the one with maximum relative error less than $10^{-6}$

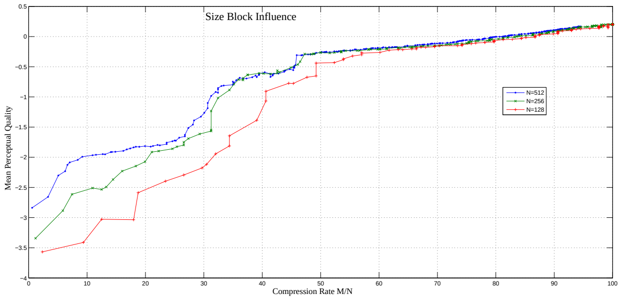

Fig. 4. PEAQ ODG of proposed system for N=128, 256 and 512 window size

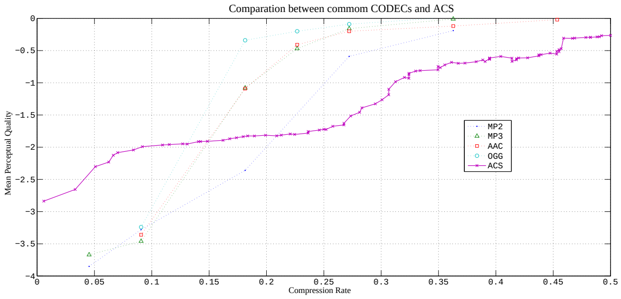

Fig. 5. Performance comparation between common CODECs and the ACS

see in figure 4 the perceptual quality of the system for different Compression rates $M/N$. The compression rate can be understood as the subsampling factor for an equivalent analog implementation of the system. We also made a comparison between our system and common CODECs like MP3, the results can be seen on figure 5. We can observe that for small compression rates (below 0.13), the system achieves better perceptual quality. Note that we propose a sampling/compression as a one step system, and not a software CODEC.

## VI. Conclusion and future work

In this paper, we evaluated the applicability of the Compressed Sensing framework to sample and compress audio signals in one step. We propose a novel technique for audio compression that has the interesting properties of error-resiliency, reverse-complexity and universality and it is the actual state-of-art of Compressed Sensing applied to audio signals.

We investigated audio sparsity basis for the intended application and found the MCDT to be the best one. Also, we evaluated reconstruction techniques for audio signals and have shown the relation between compression and sparsity basis. Finally, we evaluated the perceptual audio quality of the proposed Audio Compressed Sensing and showed that, for a given compression rate, the proposed system performs better than traditional CODECs.

The proposed technique can be applied just as it is in this paper. Nevertheless, as a natural development of this work, it can be extended to be part of a more complex audio CODEC, possibly resulting in higher compression rates. For research on sparsity basis, it can be thought as beyond transform coding like Atomic Decomposition of signals by Orthogonal Matching Pursuit [10], achieving higher levels of sparsity and thus smaller number of measurements.

## Acknowledgment

We wish to acknowledge CAPES and the PPP project (Edital 001/08-FAPESQ/MCT/CNPq) for supporting this research.

## References

[1] C. E. Shannon, “Communication in the presence of noise,” Proceedings of the Institute of Radio Engineers, vol. 37, pp. 10-21, January 1949.

[2] H. Nyquist, Certain topics in telegraph transmission theory, Transactions of the American Institute of Electrical Engineers, vol 47, pp. 617-644, January, 1928.

[3] E. J. Candès and J. R. and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” IEEE Transactions on Information Theory, vol. 52, pp. 489-509, 2006.

[4] D. L. Donoho, “Compressed sensing,” IEEE Transations on Information Theory, vol. 52, no. 4, pp. 1289-1306, April, 2006.

[5] , E. J. Candès and J. Romberg, ”Sparsity and incoherence in compressive sampling”, Inverse Problems, vol.23, pp. 969-985, 2007.

[6] J. A. Tropp and A. and C. Gilbert, ”Signal recovery from random measurements via Orthogonal Matching Pursuit”, IEEE Trans. Inform. Theory, vol. 53, pp. 4655-4666, 2007.

[7] A. Griffin and P. Tsakalides. ”Compressed Sensing of Audio Signals Using Multiple Sensors”, Proceedings of the 16th European Signal Processing Conference (EUSIPCO’08).

[8] A. Carmi and D. Kanevsky and B. Ramabhadran, ”Lossy Speech Compression Via Compressed Sensing-Based Kalman Filtering”, IBM Reseach Report, June, 2009.

[9] Thilo Thiede and William C. Treurniet and Roland Bitto and Christian Schmidmer and Thomas Sporer and John G Beerends and Catherine Colomes and Michael Keyhl and Herhard Stoll and Karlheinz Brandenburg and Bernard Feiten, ”PEAQ - the ITU standard for objective measurement of perceived audio quality”, Journal of the Audio Engineering Society, vol. 3-29, pp. 3-29, 2000.

[10] S. S, Chen , D. L. Donoho and M. A. Saunders, ”Atomic Decomposition by Basis Pursuit”, SIAM Journal on Scientific Computing, vol. 20, pp. 33-61, 1998.

[11] E. J. Candès and T. Tao, ”Near-optimal signal recovery from random projections: universal encoding strategies?”, IEEE Transactions on Information Theory, vol. 52, no 12, pp. 5406-5425, 2006.

[12] E. J. Candès , M. Rudelson , T. Tao and R. Vershynin, ”Error Correction via Linear Programming”, Annual IEEE Symposium on Foundations of Computer Science, pp. 295-308, 2005.

[13] E. J. Candès and P. A. Randall, ”Highly Robust Error Correction by Convex Programming”, IEEE Transactions on Information Theory, vol. 54, no 7, pp. 2829-2840, July, 2008.year=2008.

[14] M. Srivastava , D. Culler and D. Estrin, ”Overview of wireless sensor networks”, IEEE Computer Society August, 2004

[15] ”Sound Quality Assessment Material recordings for subjective tests”, in http://tech.ebu.ch/publications/sqamcd, acessed in april 2010.