Al-Khwarizmi Engineering Journal

Al-Khwarizmi Engineering Journal, Vol. 8, No.3, PP 53- 62 (2012)

Sub–Nyquist Frequency Efficient Audio Compression

Ahmed A. Hashim

Department of Computer/ College of Science for Women/ University of Baghdad

Email:dreng.ahmed@yahoo.com

(Received 17 April 2011; accepted 20 March 2012)

Abstract

This paper presents the application of a framework of fast and efficient compressive sampling based on the concept of random sampling of sparse Audio signal. It provides four important features. (i) It is universal with a variety of sparse signals. (ii) The number of measurements required for exact reconstruction is nearly optimal and much less then the sampling frequency and below the Nyquist frequency. (iii) It has very low complexity and fast computation. (iv) It is developed on the provable mathematical model from which we are able to quantify trade-offs among streaming capability, computation/memory requirement and quality of reconstruction of the audio signal. Compressed sensing CS is an attractive compression scheme due to its universality and lack of complexity on the sensor side. In this paper a study of applying compressed sensing on audio signals was presented. The performance of different bases and its reconstruction are investigated, as well as exploring its performance. Simulations results are present to show the efficient reconstruction of sparse audio signal. The results shows that compressed sensing can dramatically reduce the number of samples below the Nyquist rate keeping with a good PSNR.

Keywords: *Sub-Nyquist Sampling, Compressive Sampling, Compressed Sensing, Nonlinear Reconstruction, Random Matrices.*

1. Introduction

The 20th century has seen the development of a huge variety of sensors/detectors acquiring measurement in a faithful representation of the physical world (e.g. radio receivers, optical sensors, seismic detector ...). Since the purpose of these systems was to directly acquire a meaningful "signal", a very fine sampling of this latter had to be performed. This was the context surrounding the famous Shannon-Nyquist condition stating that every continuous (a priori) band-limited signal can be recovered from its discretization if its sampling rate is at least two times greater than its cutoff frequency.

Recent theory named Compressed Sensing (or Compressive Sampling) [1, 2] states that this lower bound on the sampling rate can be highly reduced, as soon as, first, the sampling is generalized to any linear measurement of the signal, and second, specific a priori hypotheses on the signal are realized. More precisely, the sensing pace is reduced to a rate that equals a few multiple of the intrinsic signal dimension rather than the dimension of the embedding space.

Technically, this simple statement is a real revolution both in the physical design of sensors and in the theory of reliable signal sampling. It means that a "given signal does not have to be acquired in its initial space as previously, but it can really be observed through a "distorting glass" (providing it is linear) with fewer measurements".

The history of Compressed Sensing has started in 2006 by the seminal works of D. Donoho, E. Cande's, T. Tao and J. Romberg [3, 4], even if some of its founding concepts, e.g. sparse recovery by convex optimization, were known from several decades. CS has actually emerged and grown from the rich multidisciplinary hotbed of Information and Sampling Theory, Inverse Problems solving, Statistics and Measure Concentration, Graph theory, and High-Dimensional (Polytope) Geometry.

In this paper I present a study of the performance of CS for a variety of audio signals and illustration the differences in performance depending on the basis and the reconstruction algorithm used.

2. Compressed Sensing

The Nyquist-Shannon sampling theorem states that to restore a signal exactly and uniquely, you need to have sampled with at least twice its frequency. Of course, this theorem is still valid; if you skip one byte in a signal or image of white noise, you cannot restore the original. But most interesting signals and images are not white noise. When represented in terms of appropriate basis functions, such as trig functions or wavelets, many signals have relatively few non-zero coefficients. In compressed (or compressive) sensing terminology, they are sparse [5].

Before starting with the mathematics related with CS let us first explain the idea with the following simple example:

Let us think of two numbers whose average is 3. What are the numbers? After complaining that there is no enough information, you might answer 2 and 4. If you do, you have unconsciously imposed a kind of regularization that requires the result to be two distinct integers; the problem is a 1–by–2 system of linear equations with matrix A=[1/2 1/2] and right–hand side b=3

We want to find a 2–vector *y* that solves *Ay=b*. The minimum norm least squares solution is computed by the pseudo inverse,
y =[3 3] but different solution is possible:x =[6 0]. Both solutions are valid, but human puzzle–solvers rarely mention them. Notice that the second solution is sparse; one of its components is zero.

The signal or image restoration problem is a larger instance of the same task; we are given thousands of weighted averages of millions of signal or pixel values. Our job is to re-generate the original signal or image.

2.1. Problem Statement of Compressible Signals

Consider a real-valued, finite-length, one-dimensional, discrete-time signal **x**, which can be viewed as an $N \times 1$ column vector in $\mathbb{R}^N$ with elements $x[n]$, $n = 1, 2, \ldots, N$. Any signal in $\mathbb{R}^N$ can be represented in terms of a basis of $N \times 1$ vectors $\{\psi_i\}_{i=1}^{N}$. Using the $N \times N$ basis matrix $\psi = [\psi_1|\psi_2| \ldots |\psi_N]$ with the vectors $\{\psi i\}$ as columns, a signal **x** can be expressed as

$$\mathbf{X} = \sum_{i=1}^{N} s_i \psi_i \quad \text{or} \quad \mathbf{X} = \mathbf{s}\boldsymbol{\psi} \qquad \ldots(1)$$

Where **s** is the $N \times 1$ column vector of weighting coefficients $si = \langle x, \psi_i \rangle = \psi_i^T$ **x**. Clearly, **x** and **s** are equivalent representations of the signal, with **x** in the time or space domain and **s** in the ψ domain. The signal **x** is *K-sparse* if it is a linear combination of only *K* basis vectors; that is, only *K* of the *si* coefficients in (1) are nonzero and $(N - K)$ are zero. The case of interest is when $K \ll N$. The signal **x** is *compressible* if the representation (1) has just a few large coefficients and many small coefficients.

2.2. Transform Coding and its Inefficiencies

The fact that compressible signals are well approximated by *K*-sparse representations forms the foundation of transform coding [3, 6]. In data acquisition systems (for example, digital cameras) transform coding plays a central role: the full *N*-sample signal **x** is acquired; the complete set of transform coefficients $\{si\}$ is computed via $\mathbf{s} = \psi^T\mathbf{x}$; the *K* largest coefficients are located and the $(N - K)$ smallest coefficients are discarded; and the *K* values and locations of the largest coefficients are encoded. Unfortunately, this sample–then–compress framework suffers from three inherent inefficiencies. First, the initial number of samples *N* may be large even if the desired *K* is small. Second, the set of all *N* transform coefficients $\{si\}$ must be computed even though all but *K* of them will be discarded. Third, the locations of the large coefficients must be encoded, thus introducing an overhead.

2.3. The Compressive Sensing Problem

Compressive sensing address these inefficiencies by directly acquiring a compressed signal representation without going through the intermediate stage of acquiring *N* samples [1, 7, 8]. Consider a general linear measurement process that computes $M < N$ inner products between **x** and a collection of vectors $\{\phi_j\}_{j=1}^{M}$ as in $y_j = \langle \mathbf{x}, \phi_j \rangle$. Arrange the measurements y_j in an $M \times 1$ vector **y** and the measurement vectors ϕ_j^T as rows in an $M \times N$ matrix Φ. Then, by substituting Ψ from (1), **y** can be written as

$$y = \Phi X = \Phi \Psi S = \Theta S \qquad \ldots(2)$$

where $\Theta = \Phi\Psi$ is an $M \times N$ matrix. The measurement process is not adaptive, meaning that Φ is fixed and does not depend on the signal **x**. The problem consists of designing ***a)*** a *stable measurement matrix* Φ such that the salient information in any *K*-sparse or compressible signal is not damaged by the dimensionality reduction from $x \in \mathbb{R}^N$ to $y \in \mathbb{R}^M$ and ***b)*** a *reconstruction algorithm* to recover **x** from only $M \approx K$ measurements **y** (or about as many measurements as the number of coefficients recorded by a traditional transform coder).

2.4. Designing a Stable Measurement Matrix

The measurement matrix Φ must allow the reconstruction of the length-*N* signal **x** from $M < N$ measurements (the vector **y**). Since $M < N$, this problem appears ill-conditioned. If, however, **x** is *K*-sparse and the *K* locations of the nonzero coefficients in **s** are known, then the problem can be solved provided $M \geq K$. A necessary and sufficient condition for this simplified problem to be well conditioned is that, for any vector **v** sharing the same *K* nonzero entries as **s** and for some $\epsilon > 0$

$$1 - \epsilon \leq \frac{\|\Theta v\|_2}{\|v\|_2} \leq 1 + \epsilon \qquad \ldots(3)$$

That is, the matrix Θ must preserve the lengths of these particular *K*-sparse vectors. Of course, in general the locations of the *K* nonzero entries in **s** are not known. However, a sufficient condition for a stable solution for both *K*-sparse and compressible signals is that Θ satisfies (3) for an arbitrary 3*K*-sparse vector **v**. This condition is referred to as the *restricted isometry property* (RIP) [4]. A related condition, referred to as *incoherence*, requires that the rows $\{\phi_j\}$ of Φ cannot sparsely represent the columns $\{\psi_i\}$ of ψ (and vice versa). Direct construction of a measurement matrix Φ such as $\Theta = \Phi\Psi$ has the RIP that requires verifying (3) for each of the $\binom{N}{K}$ possible combinations of *K* nonzero entries in the vector **v** of length *N*. However, both the RIP and incoherence can be achieved with high probability simply by selecting Φ as a random matrix. For instance, let the matrix elements $\phi_{j,i}$ be independent and identically distributed (iid) random variables from a Gaussian probability density function with mean zero and variance 1/*N* [1, 2, 4]. Then the measurements **y** are merely *M* different randomly weighted linear combinations of the elements of **x**, as illustrated in Fig. 1(a).

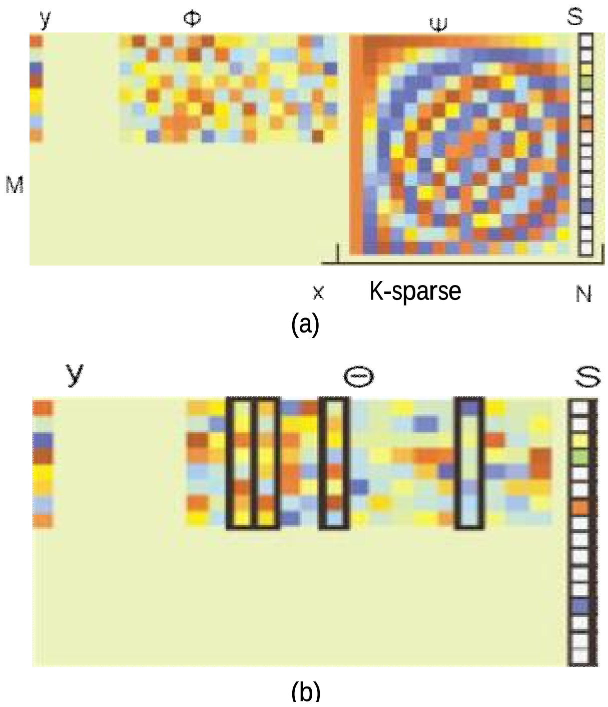


Fig. 1. (a) Compressive Sensing Measurement Process with a Random Gaussian Measurement Matrix *and* Discrete Cosine Transform (DCT) Matrix. The Vector of Coefficients s is Sparse with K = 4. (b) Measurement Process with $\Theta = \Phi\Psi$ There are Four Columns that Correspond to Nonzero s_i Coefficients; the Measurement Vector y is a Linear Combination of These Columns [9].

The Gaussiam measurement matrix has two interesting and useful properties:

a- The matrix Φ is incoherent with the basis Ψ = **I** of delta spikes with high probability. More specifically, an $M \times N$ iid Gaussian matrix Θ = Φ **I** = Φ can be shown to have the RIP with high probability if ***M* ≥ *c K* log (*N/K*),** with *c* a small constant [1, 2, 4]. Therefore, *K*-sparse and compressible signals of length *N* can be recovered from only ***M* ≥ *cK* log(*N/K*)**$\ll N$ random Gaussian measurements.

b- The matrix Φ is universal in the sense that $\Theta = \Phi\Psi$ will be iid Gaussian and thus have RIP with high probability regardless of the choice of orthonormal basis Ψ .

2.5. Designing a Signal Reconstruction Algorithm

The signal reconstruction algorithm must take the *M* measurements in the vector **y**, the random measurement matrix Φ (or the random seed that generated it), and the basis Ψ and reconstruct the length–*N* signal **x** or, equivalently, its sparse coefficient vector **s**. For *K*-sparse signals, since *M* < *N* in (2) there are infinitely many $\hat{s}$ that satisfy $\Theta\ \hat{s}$ = **y**. This is because if Θ **s** = **y** then Θ *(***s** + **r***)* = **y** for any vector **r** in the null space $N(\Theta)$ of Θ. Therefore, the signal reconstruction algorithm aims to find the signal's sparse coefficient vector in the *(N − M)*-dimensional translated null space $\mathcal{H} = N(\Theta) + s$.

a- Minimum *l2* norm reconstruction:

Define the *lp* norm of the vector **s** as

$$[(\|s\|]_p)^p = \sum_{i=1}^{N} |s_i|^p$$

. The classical approach to inverse problems of this type is to find the vector in the translated null space with the smallest *l*2 norm (energy) by solving

$$\hat{s} = \text{argmin } \|s'\|_2 \quad \text{such that } \Theta s' = \mathbf{y} \qquad \text{...(4)}$$

This optimization has the convenient closed-form solution $\hat{s} = \Theta^T(\Theta\Theta^T)^{-1} y$.
Unfortunately, *l*2 minimization will almost never find a *K*-sparse solution, returning instead a non sparse $\hat{s}$ with many nonzero elements.

b- Minimum l0 norm reconstruction:

Since the *l*2 norm measures signal energy and not signal sparsity, consider the *l*0 norm that counts the number of non-zero entries in **s**. (Hence a *K*-sparse vector has *l*0 norm equal to *K*). The modified optimization

$$\hat{s} = \text{argmin } \|s'\|_0 \quad \text{such that } \Theta s' = \mathbf{y} \qquad \text{...(5)}$$

can recover a *K*-sparse signal exactly with high probability using only *M* = *K* + 1 iid Gaussian measurements [5]. Unfortunately, solving (5) is both numerically unstable and NP complete, requiring an exhaustive enumeration of all $\binom{N}{K}$ possible locations of the nonzero entries in **s**.

c- Minimum l1 norm reconstruction:

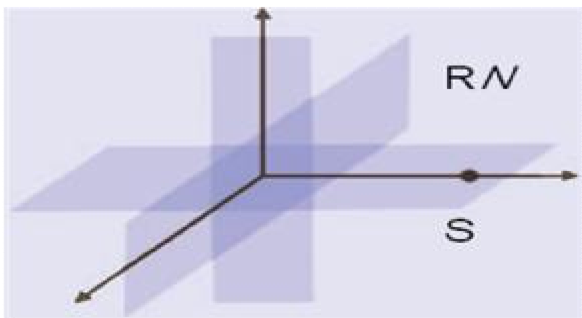
Surprisingly, optimization based on the *l*1 norm

$$\hat{s} = \text{argmin } \|s'\|_1 \quad \text{such that } \Theta s' = \mathbf{y} \qquad \text{...(6)}$$

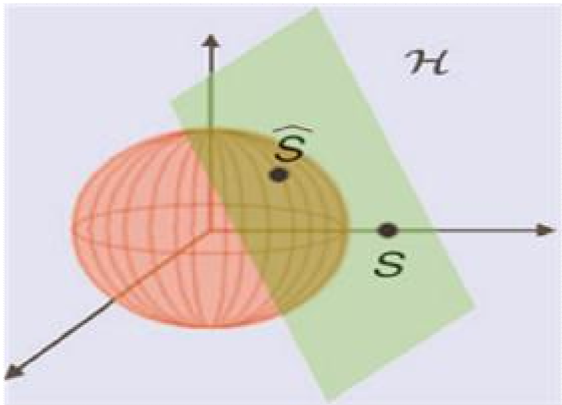
can exactly recover K-sparse signals and closely approximate compressible signals with high probability using only **M ≥ cK log(N/K)** iid Gaussian measurements [1], [2]. This is a convex optimization problem that conveniently reduces to a linear program known as basis pursuit whose computational complexity is about $O(N^3)$.

2.6. The Reason Behind the Convergence of l1 rather than l2

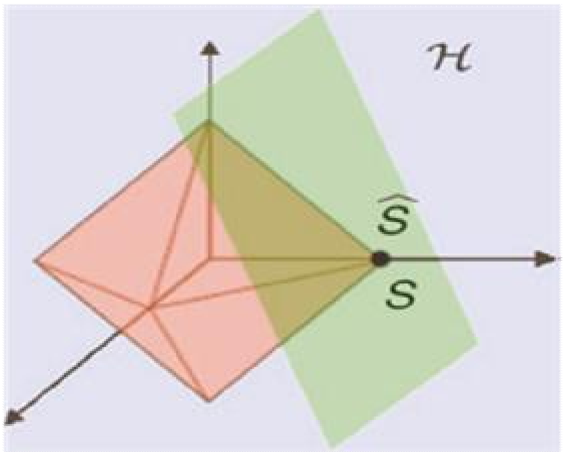
The geometry of the compressive sensing problem in $\mathbb{R}^N$ helps visualize why *l*2 reconstruction fails to find the sparse solution that can be identified by *l*1 reconstruction. The set of all *K*-sparse vectors **s** in $\mathbb{R}^N$ is a highly nonlinear space consisting of all *K*-dimensional hyper planes that are aligned with the coordinate axes as shown in Fig. 2(a). The translated null space $\mathcal{H} = N(\Theta) + s$ is oriented at a random angle due to the randomness in the matrix Θ as shown in Fig. 2(b). (In practice *N*, *M*, *K* ≫ 3, so any intuition based on three dimensions may be misleading) The *l*2 minimizer $\hat{s}$ from (4) is the point on $\mathcal{H}$ closest to the origin. This point can be found by blowing up a hyper sphere (the *l*2 ball) until it contacts $\mathcal{H}$. Due to the random orientation of $\mathcal{H}$, the closest point $\hat{s}$ will live away from the coordinate axes with high probability and hence will be neither sparse nor close to the correct answer **s**. In contrast, the *l*1 ball in Fig. 2(c) has points aligned with the coordinate axes. Therefore, when the *l*1 ball is blown up, it will first contact the translated null space $\mathcal{H}$ at a point near the coordinate axes, which is precisely where the sparse vector **s** is located. While the focus here has been on discrete- time signals **x**.



a



b

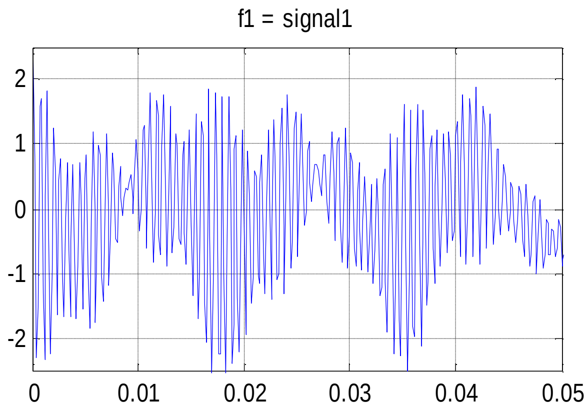


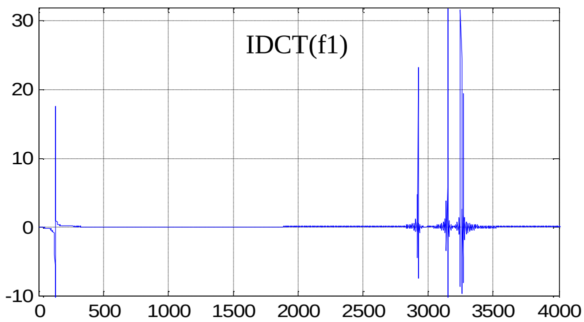
c

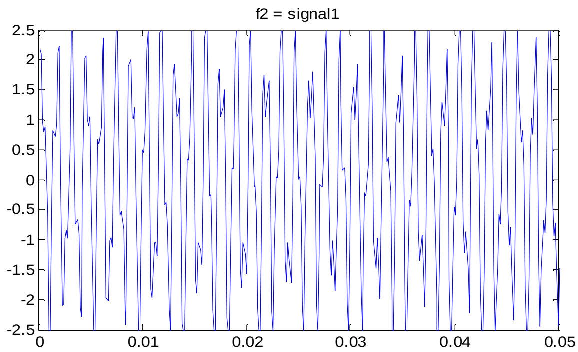
Fig. 2. (a) Subspaces with two Sparse Vectors in R3 lie Close to the Coordinate Axes. (b) Visualization of the *l2* Minimization (5) that Finds the Non-Sparse Point-of-Contact *s* between the 2 Ball (Hyper-Sphere, in Red) and the Translated Measurement Matrix Null Space (in Green). (c) Visualization of the *l1* Minimization Solution that Finds the Sparse Point-of-Contact *s* with High Probability Thanks to the Pointiness of the *l1* ball.

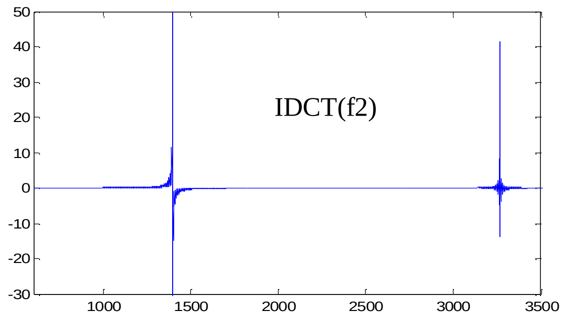
3. Simulation Results

Three types of signals are taken, based on complexity in time domain and in terms of sparsity, see Fig. 3. These signals are sampled and then reconstructed from few randomly selected samples, Fig. 4 shows the sampling of the first signal of cutoff frequency 1.633kHz with sampling frequency of 14kHz and then taking randomly 10% of these samples to reconstruct the signal. The figure shows that reconstruction with *l1* – norm is accurate with PSNR of 20.5dB while reconstructing using *l2* – norm was very pad and gave meaningless results. Fig. 5 shows the original signal3 (with highest sparsity) with cutoff frequency of 1633 and the reconstructed one with different random samples (m)/total samples (n) rations.









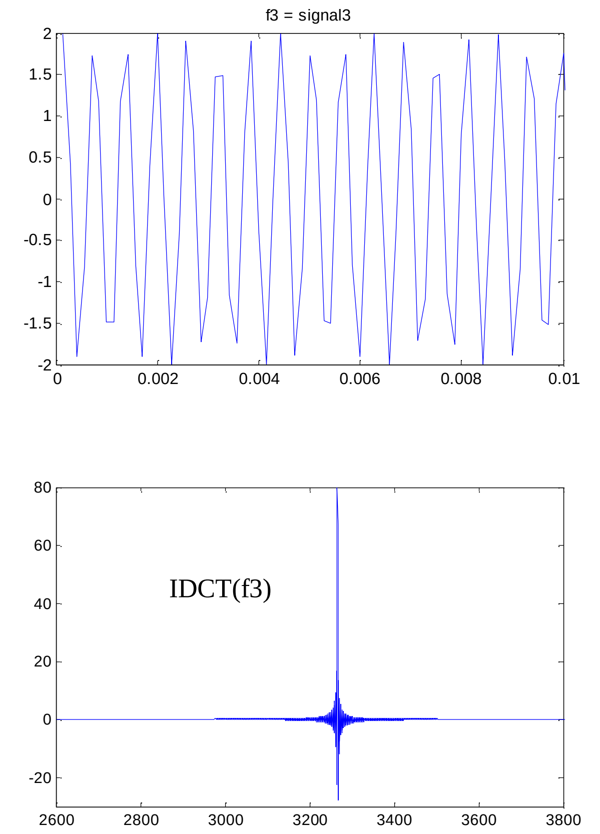
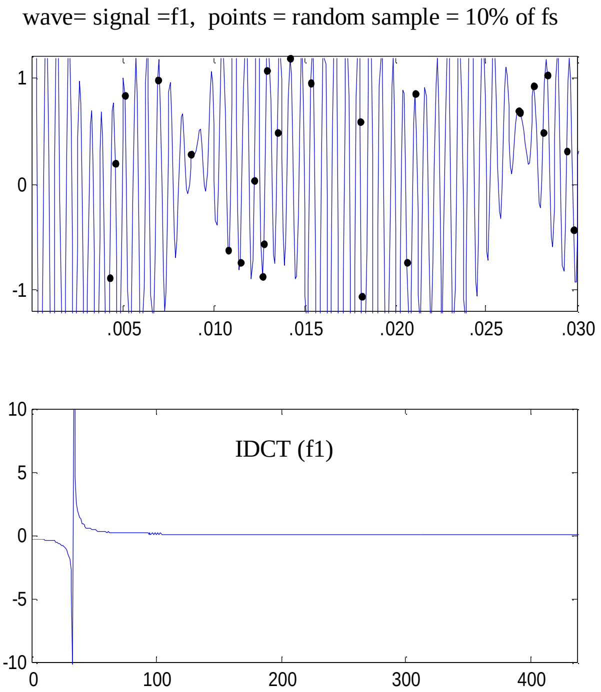


Fig. 3. Three Time Domain Signals with their IDCT.



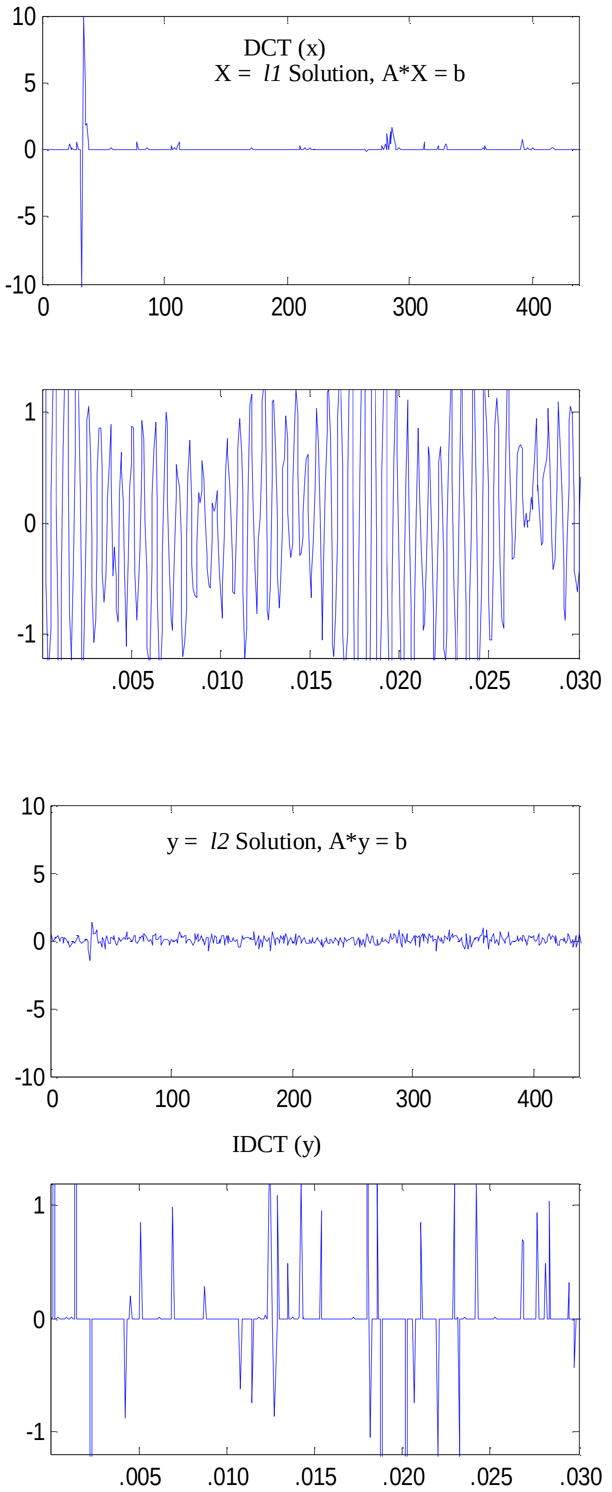
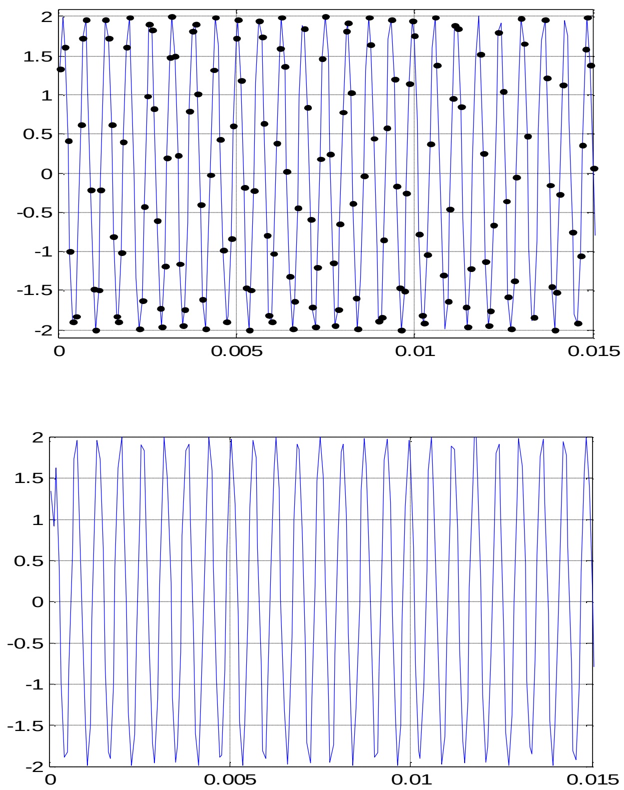
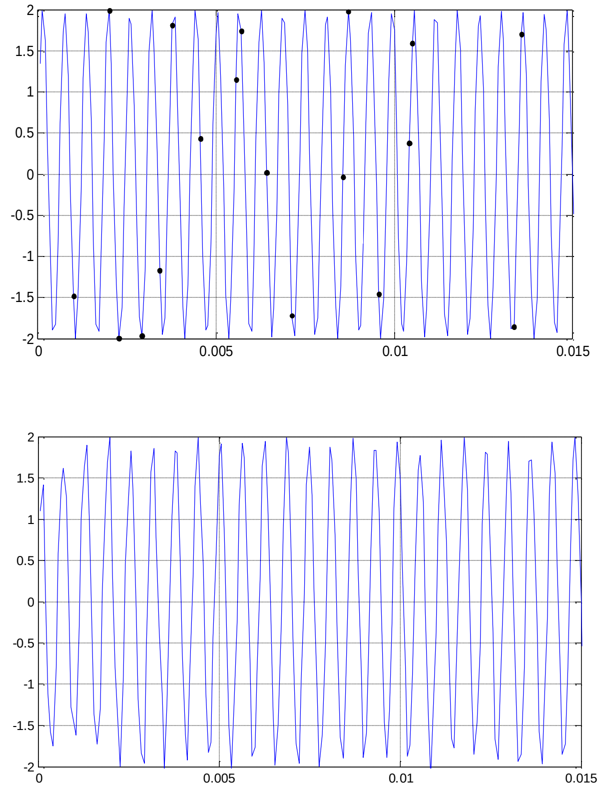


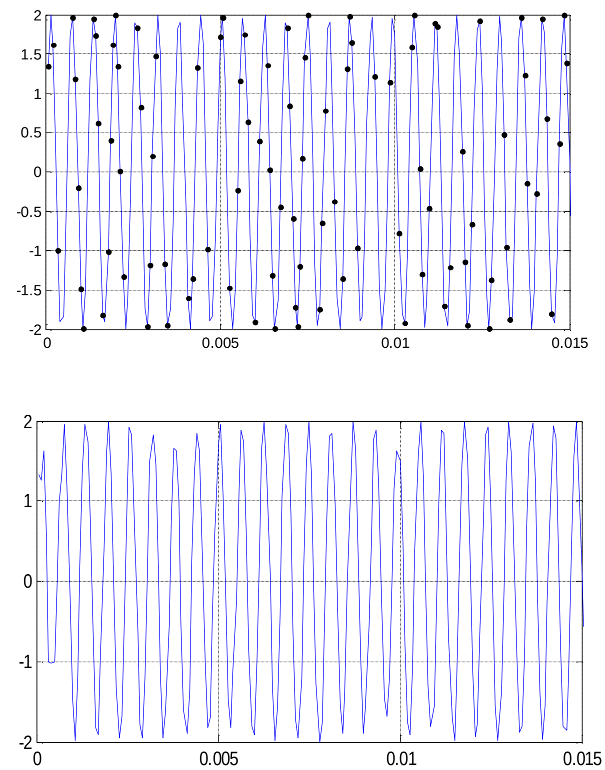
Fig. 4. a) Signal and its Random Samples b) Its sparse representation in idct c) Reconstruction Using *l1* in the Sparse Domain d) Reconstructed Signal Using l1 e) Reconstruction Using *l2* in the Sparse Domain f) Reconstructed Signal.



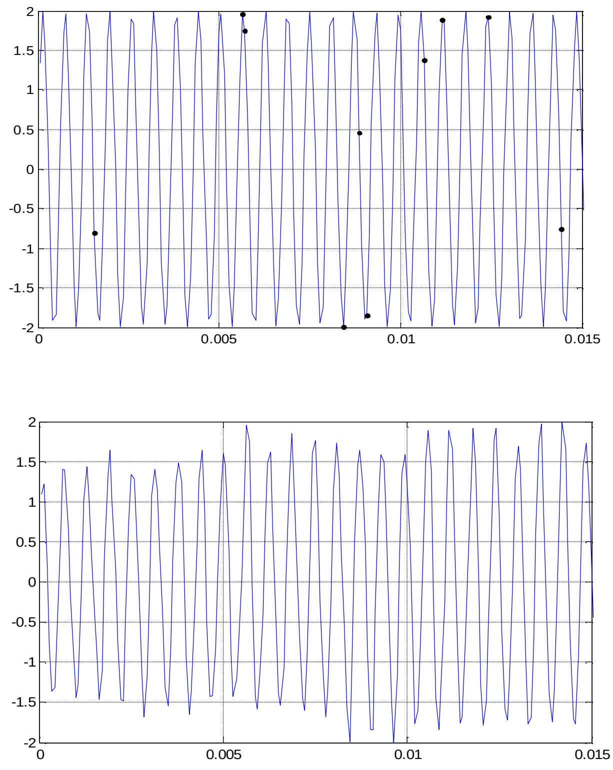
(a) Original and Reconstructed Signal with m/n =0.8, PSNR = 38.6.



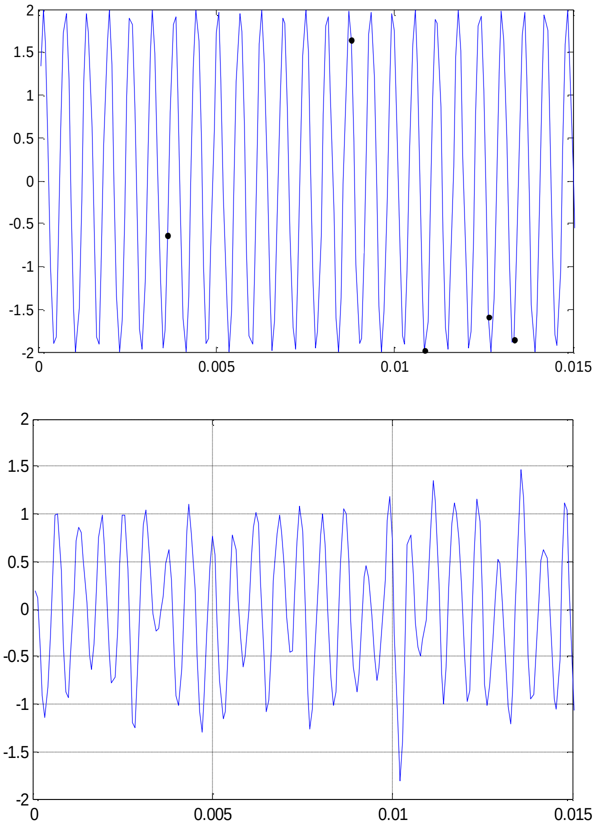
(c) Original and Reconstructed Signal with m/n = 0.1, PSNR = 21.



(b) Original and Reconstructed Signal with m/n = 0.5, PSNR = 32.



(d) Original and Reconstructed Signal with m/n =0.05, PSNR = 19.6.



e) Original and reconstructed signal with m/n =0.02, PSNR = 12.5

Fig. 5. Original and Reconstructed Signal with Different m/n Ratios and the PSNR for them, the Cutoff Frequency of the Signal is 1.63kHz and Fs = 14kHz

From Fig.5, one can see the good reconstruction even when the m/n ratio is small; the PSNR for each case reflects the goodness of reconstruction.

Fig. 6 shows the PSNR versus m/n ratio for the three under testing signals, from the Figure one can see that as the sparsity of the signal increases; the reconstruction with lower m/n ratio is possible. It is important here to say that since the reconstruction process is based on random sampling; the PSNR gained may vary based on (by chance) hitting the target (the random samples takes the largest values of the signal in the sparse domain)

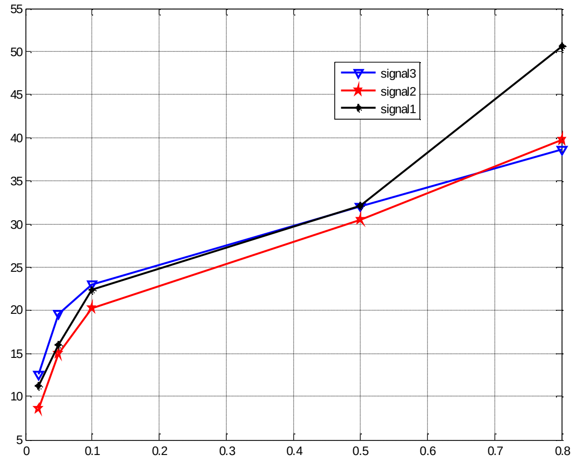


Fig. 6. The PSNR Versus m/n Ratio for the Three Under Testing Signals.

It is convenient here to say that compressive sensing also applies to sparse or compressible analog signals *x(t)* as well as digital ones

4. Conclusions

Signal acquisition based on compressive sensing can be more efficient than traditional sampling for sparse or compressible signals. In compressive sensing, the familiar least squares optimization is inadequate for signal reconstruction, and other types of convex optimization must be invoked. The CS is Nonlinear sampling, so that it is an arbitrary and unknown set of size K, exact recovered from cKLog(N/K) (almost) arbitrarily placed samples, and nonlinear reconstruction by convex programming.

It is important to mention here that the MP3 and JPEG files used by today's audio systems and digital cameras are already compressed in such a way that exact reconstruction of the original signals and images is impossible.

5. References

[1] Justin Romberg and Michael Wakin "Compressed Sensing: A Tutorial", IEEE Statistical Signal Processing Workshop, Madison, Wisconsin, 2007.

[2] Albert Cohen, et al. "Compressed sensing and best k-term approximation" , Naval Resarch Contracts ONR-N00014-03-1-0051, 2006

[3] Yin Zhang " On Theory of Compressive Sensing via l1-Minimization: Simple Derivations and Extensions", CAAM Technical Report TR08-11, 2008

[4] Emmanuel Cand`es and Justin Romberg " l1-magic: Recovery of Sparse Signals via Convex Programming", 2005

[5] Alyson K. Fletcher "Necessary and Sufficient Conditions for Sparsity Pattern Recovery", IEEE, revision of arXiv:0804.1839v1 [cs.IT], 2009

[6] Cleves Corner "the World's Simplest Impossible Problem", The MathWorks Newsletter, Vol.4, No. 2, 1990

[7] P. Wojtaszczyk "Stability of l1 minimisation in compressed sensing",2009

[8] Yin Zhang "When is missing data recoverable?", CAAM Technical Report TR06-15, 2006

[9] Kaichun K. Chang , …etal, “Sub-Nyquist Audio Fingerprinting for music Recognition”, 2010

ضغط صوت كفوء بإستعمال تردد أقل من نـايكويست

أحمد عبدالصاحب هاشم
قسم علوم الحاسوب/ كلية العلوم للبنات/ جامعة بغداد
البريد الالكتروني: dreng.ahmed@yahoo.com

الخلاصة

تعرض هذه الورقة تطبيق إطار ضغط سريع وفعال لأخذ العينات على أساس مفهوم العينة عشوائية من إشارة الصوت المتناثر. هذا النوع من الضغط يوفر أربع سمات هامة. (أ) هي طريقه شاملة لمجموعة متنوعة من الإشارات المتناثرة (ب) عدد القياسات المطلوبة لإعادة تشكيل الاشاره الصوتية بالضبط هو الأمثل تقريبا وأقل بكثير من التردد المستعمل لأخذ العينات وأقل من تردد نايكويست (ت) منخفضة التعقيد للغاية وسرعه بالحسابات (ث) من خلال وضعها في نموذج رياضي يمكن اثباته نحن قادرون على تحديد مقارنة بين قدرة حساب حجم تدفق المبادلات نسبه الى الذاكرة المطلوبة، ونوعية إعادة البناء للإشارات الصوتية. الاستشعار المضغوط CS هو إسلوب ضغط جذاب نظراً لعالميته وعدم وجود تعقيد في جانب الاستشعار. في هذه الورقة تم تقديم دراسة لتطبيق الاستشعار المضغوط على الإشارات الصوتية. تم التحقيق من أداء الطريقه وقدرتها على اعادة تشكيل الاشارة بتطبيقها على اشارات ذات اسس مختلفه ، وكذلك استكشاف أداءها. نتائج المحاكاة موجودة لإظهار كفاءة إعادة تشكيل الإشارات الصوتية المتناثرة. النتائج تظهر أن الاستشعار المضغوط يمكن أن يحد بشكل كبير من عدد العينات المطلوبه لأعادة تشكيل الاشارة الصوتية أقل من معدل نايكويست مع المحافظة على نسبة أعلى اشارة الى ضوضاء PSNR جيدة.