DCT-compressive Sampling of Frequency-sparse Audio Signals

R.G. Moreno-Alvarado, Mauricio Martinez-Garcia

***Abstract*—The discrete cosine transform (DCT) and the compressive sampling (CS) are two signal processing techniques with many applications on a great number of engineering fields. In this paper, we propose to apply both techniques to the compression of audio signals.**

Using spectral analysis and the properties of the DCT, we can treat audio signals as sparse signals in the frequency domain. This is especially true for sounds representing tones. On the other hand, CS has been traditionally used to acquire and compress certain sparse images. We propose the use of DCT and CS to obtain an efficient representation of audio signals, especially when they are sparse in the frequency domain.

By using the DCT as signal preprocessor in order to obtain a sparse representation in the frequency domain, we show that the subsequent application of CS represent our signals with less information than the well-known sampling theorem. This means that our results could be the basis for a new compression method for audio and speech signals.

***Index Terms*—Audio signals, Compressive sampling, DCT, sparsity.**

I. Introduction

Over the past few years, there has been an increased interest in the study of compressed sampling (CS),a new framework for sampling and compressing certain signals. In CS, the bandlimited model (i.e. the Nyquist sampling theorem) is replaced by a sparse model, assuming that a signal can be efficiently represented using only a few significant coefficients in some transform domain.

The groundbreaking work by Candes et al. [1] and Donoho [2] showed that such a signal can be precisely reconstructed from only a small set of random linear measurements (smaller than the Nyquist rate), implying the potential of dramatic reduction of sampling rates, power consumption and computation complexity in digital data acquisitions.

Manuscript received March 23, 2011; revised April 13, 2011. This work was supported in part by Universidad La Salle, Mexico, under project I-48/09.
R.G. Moreno-Alvarado iswith IPN Escuela Superior de IngenieriaMecanica y ElectricaCulhuacan,Seccion de Estudios de Posgrado e InvestigacionMexico D.F.,E-mail: rmoreno@calmecac.esimecu.ipn.mx
Mauricio Martinez-GarciaiswithUniversidad la Salle, Facultad de Ingenieria,Benjamin Franklin 47, C.P. 06140, Mexico D.F., Email: mauricio.martinez@lasallistas.org.mx

Due to the large amount of data in image signals, CS is very attractive in imaging applications, especially for low-power and low resolution imaging devices or when the measurement is very costly (e.g., Terahertz applications). Since the discovery of the CS theory, several compressive imaging algorithms have been developed for Fourier transform domain measurements in applications such as the MRI [3].

Despite the above mentioned work, there still exists a huge gap between the CS theory and applications to audio signals [11], [12]. In particular, it is still unknown how to construct a sparse audio signal, especially when CS relies on two principles: sparsity (which pertains to the signal of interest), and incoherence (which pertains to the sensing modality) [4-6].

For the problem of making asparse representation of an audio signal, we introduce the DCT which is at present, the most widely used transform for image and video compression systems. Its popularity is due mainly to the fact that it achieves a good data compaction, because it concentrates the information content in relatively few coefficients[7]. This means that we can obtain a compressed version of an audio signal by first obtaining a sparse representation in the frequency domain, and later processing the result with aCS algorithm.

II. Background

A. Compressive Sampling

A recent series of papers [1–6] develop a theory of signal recovery from highly incomplete information. The central results state that a sparse vector $x^0 \in R^N$be recovered from a small number of linear measurements$b = Ax^0 \in R^N, K \ll Norb = Ax_0 + e$,when e is the measurement noise by solving a convex program.

Consider a length$N$, real valued signal x and suppose that the basis ψ provides a K sparse representation of x. In terms of matrix notation, we have $x = \psi f$ in which f can be well approximated using only $K << N$non zero entries and ψ is called as the sparse basis matrix[2].

The CS theory states that such a signal x can be reconstructed by taking only $M = O(KlogN)$ linear, non adaptive measurements as follows[1,2]:

$$y = \phi x = \phi\psi f \cdots (1)$$

Where y represents an$M \times 1$ sampled vector and$\phi$is an $M \times N$ measurement matrix that is incoherent withψ, ie.,the maximum magnitude of the element in $\phi\psi$is small [5].

Finally, with this information we decide to recover the signal by ℓ_1 norm. When f is sufficiently sparse, the recovery via ℓ_1-minimization is probably exact [1].

B. The One-Dimensional DCT

The most common DCT definition of a 1-D sequence of length *N* is:

$$c(u) = \propto (u) \sum_{X=0}^{N-1} f(x) cos \left[\frac{\pi(2x+1)u}{2N} \right] \cdots (2)$$

for $u = 0,1,2, \ldots, N-1$. Similarly, the inverse transform is defined as:

$$f(x) = \propto (u) \sum_{X=0}^{N-1} \alpha(u) c(u) cos \left[\frac{\pi(2x+1)u}{2N} \right] \cdots (3)$$

for $x = 0,1,2, \ldots, N-1$. In both equations (2) and (3) $\alpha(u)$ is defined as:

$$\alpha(u) = \sqrt{\frac{1}{N}} \text{ for } u = 0 \alpha(u) = \sqrt{\frac{2}{N}} \text{for } u \neq 0$$

It is clear from (3)that for $u = 0, c(u) = \sqrt{\frac{1}{N}} \sum_{X=0}^{N-1} f(x)$

Thus, the first transform coefficient isthe average value of the sample sequence. In the literature, this value is referred to as the *DCCoefficient*. All other transform coefficients are called the *AC Coefficients*[7].

C. Properties of DCT

Decorrelation

Themain advantage of signal transformation is the removal ofredundancy between neighboring values. This leads to uncorrelated transform coefficients whichcan be encoded independently.

Energy Compaction

Efficacy of a transformation scheme can be directly gauged by its ability to pack input data intoas few coefficients as possible. This allows the quantizer to discard coefficients with relativelysmall amplitudes without introducing visual distortion in the reconstructed image. DCT exhibits excellent energy compaction for highly correlated signals.

III. PROPOSED IMPLEMENTATION OF DCT AND COMPRESSIVE SAMPLING

In this section, we introduce our proposed techniques applied to an audio signal, and describe the technique for representing it in a sparse way. We then analyze its application to a compressive sampling algorithm.

Simply speaking, we consider the aspect recovering sparse signals from just a few measurements. In this case of study, compressive sampling needs to deal with speech signals which are only approximately sparse.

The issue here is to obtain an accurate reconstruction of such signals from highly undersampled measurements.Ideally, we would like to measure all the N coefficients of f, but in the CS framework we only get to observe a subset of these and collect the data.

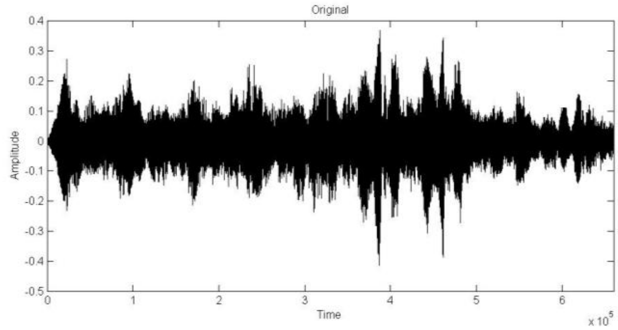

Fig. 1 Music Signal

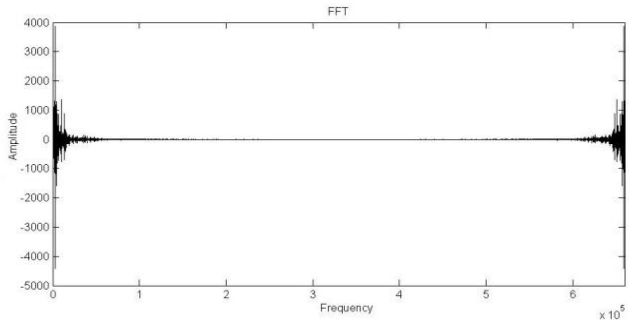

Fig. 2 FFT amplitude of Music Signal

As we can observe in figure (1), our audio signal is not sparse in the time domain. In spite of this, we applied as first instance, the fast Fourier transform (FFT) for obtaining the frequency domain representation. We can observe that this representation takes the form of a sparse signal (figure 2).

However, this representation has real and complex parts, which result in a difficult reconstruction due to the phase angle changes with the matrix transformations on the compressive sampling program, as described in [8]. This means that we cannot get the original signal by just applying the inverse Fast Fourier Transform.

As a second experiment, we used a special case of the FFT called the DCT [7], [9], [10]. As mentioned above, one of the properties is that it attempts to decorrelate the data. After decorrelation, each transform coefficient can be encoded independently without losing compression efficiency.

IV. EXPERIMENTAL RESULTS

In this sectionwe present the use of the DCT to preprocess a 15 second music piece, in order to obtain a sparse representation of our signal in the frequency domain. Subsequent application of a CS algorithm for compression and recovery succeeds in obtaining an intelligible signal.

Let us consider a stationary 15-second music piece with length of 661000 samples subdivided into 661 frames of 1000 samples each one, to increase the CS algorithm accuracy [8]. By applying the DCT for each one, we obtain

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

a sparse representation of the music signal frames (figure 3), which is considered optimal for our experiment.

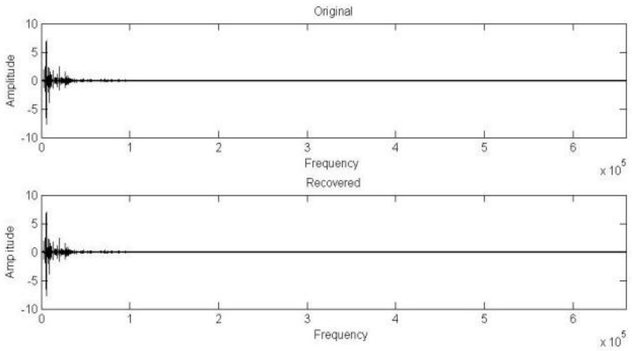

Fig. 3 First row DCT signal, second row DCT recovered signal

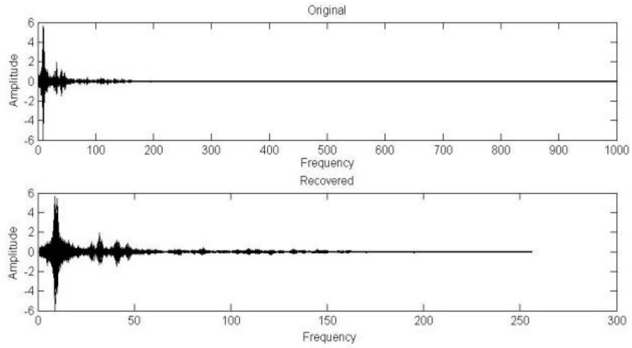

Fig. 4 First row DCT signal, second row DCT recovered signal with different length

Considering these frames of music signals as sparse, we can apply the algorithm of CS to each frame with length of 1000 samples, as a first experiment. As a second experiment, we apply the algorithm of CS to each frame with different DCT-lengths as 128,256,512,700 DCT-length of 256 is shown in figure (4).

The performance of the CS has been evaluated using the following reconstruction algorithm and software: Min- with equality constraints solver in the magic package [8].

As a result of applying the CS algorithm, we obtain a music signal nearly equal as the input sparse signal (figures 3 and 4). For the special case of experiment II, we have to apply the inverse discrete cosine transform (IDCT) to convert the data back to time domain, and recover the original speech signal (figures 5 and 6).

Tables (I, II) summarize the results of our experiments, for both DCT with length of 1,000 and DCT with different lengths of a music signal, and its reconstruction similarities. We highlighted the best result in bold letters. Figures (7) and (8) show the graphic similarities for DCT, and figure (9) show the graphic similarities per 1000-sample frames.

As shown in the figures (7) and (8) we see that as more observations are made, the similarity of both original DCT and DCT recovered signal increases. This property depends on the sparsity of the signals: a signal with a highly sparse representation decreases the number of samples.

For these experiments we have to consider a number of samples, and the similarity measures given by:

$$Similarity = e^2/E^2 \cdots (4)$$

Where e^2 is the matrix error given by the norm of $\|x - x_p\|$ divided by his length, and E^2 is the matrix power given by the norm of $\|x\|$ divided by his length.

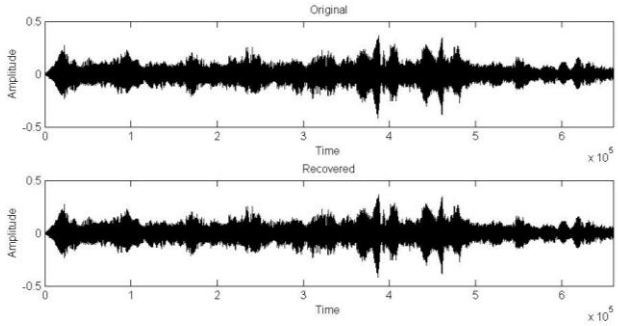

Fig. 5 first row original signal, second row recovered signal

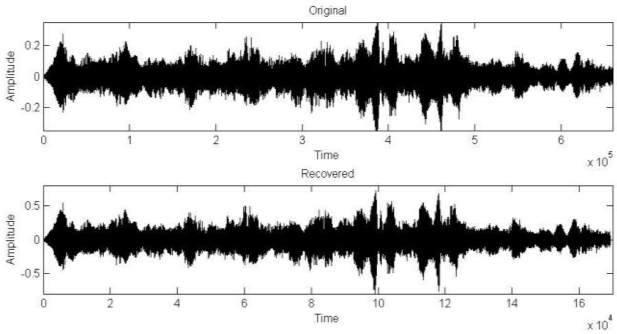

Fig. 6 First row original signal, second row recovered signal with different length

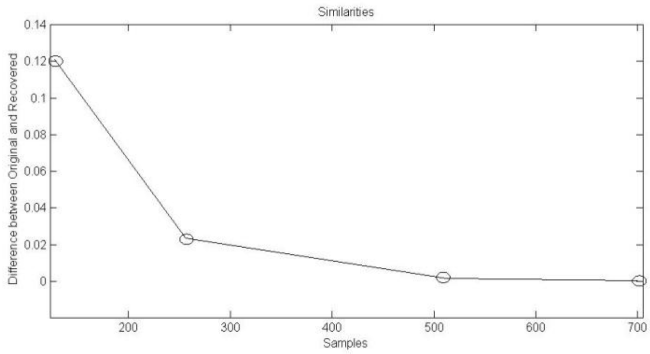

Fig. 7 Similarities with DCT frame of 1,000

TABLE I
SUMMARY OF SIMILARITIES IN MUSIC EXPERIMENT I

Samples	DCT Frame	Similarity
128	1,000	0.1208
256	1,000	0.0234
512	1,000	0.0014
700	**1,000**	**3.27e-04**

TABLE II
SUMMARY OF SIMILARITIES IN MUSIC EXPERIMENT II

Samples	DCT Frame	Similarity
164	128	3,11e-06
228	256	0.0012
356	512	0.0017
450	**700**	**0.0012**

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

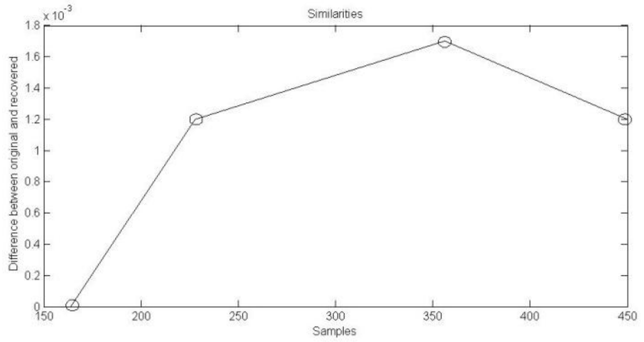

Fig. 8 Similarities with different DCT frames

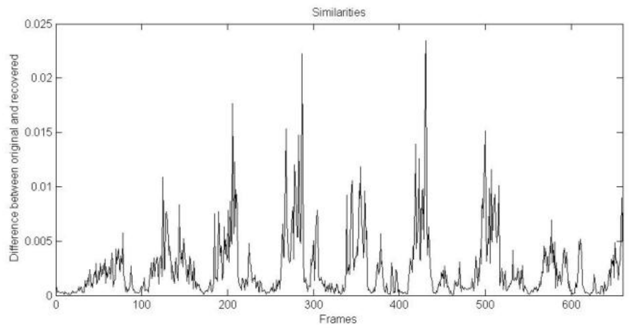

Fig. 9 Similarities per frames of 1,000 samples

V. Conclusion

This paper has proposed an efficient joint implementation of DCT, as a method to obtain a sparse audio signal representation, and the application of the compressive sampling algorithm to this sparse signal. The music is the Bach-Cello suite No. 1 Prelude.

The DCT speech signal representation has the ability to pack input data into as few coefficients as possible. This allows the quantizer to discard coefficients with relatively small amplitudes without introducing audio distortion in the reconstructed signal.

Although the compressive sampling technique is used primarily for compression sample images, we achieve reasonable results due to the preprocessing of the audio signal.

This means that our hypothesis is satisfied in the sense that our proposed technique can achieve a significant reduction in the number of samples required to represent certain audio signals, and therefore a decrease of the required number of bytes for encoding.

It was found that the audio compression model proposed in this paper is feasible, and can achieve significant compression of the music signal that can reach a value in some cases about 50% of compression with a reasonable quality depending on the particular application.

The compression values obtained are varied and depend largely on how sparse signal can be arranged and the level of quality that you want, but for a reasonable quality music must have [DCT/2+100] observations.

References

[1] E. Candes, J. Romberg, and T. Tao, "*Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*,"IEEE Trans. Inform. Theory, vol. 52, pp.489–509, Feb. 2006.
[2] D. L. Donoho, "*Compressed sensing*," IEEE Trans. Inform. Theory, vol. 52, pp. 1289–1306, July 2006.
[3] M. Lustig, D.L. Donoho, J. Santos, and J. Pauly, "*Compressed sensing MRI*," IEEE Signal Processing Magazine, March 2008.
[4] E. Candes and J. Romberg, "*Robust signal recovery from incomplete observations*," in Proc. ICIP, 2006, pp. 1281–1284.
[5] E. Candes and J. Romberg, "*Sparsity and incoherence in compressive sampling*," Inverse Problems, vol. 23, pp. 969–985,June 2007.
[6] E. Candes and J. Romberg, "*Practical signal recovery from random projections*," 2005,preprint. [Online]. Available: www.acm.caltech.edu/emmanuel/papers/PracticalRecovery.pdf
[7] Sayed Ali Khayam, "*The Discrete Cosine Transform (DCT):Theory and Application*" Department of Electrical & Computer Engineering Michigan State University, 2003
[8] l1-magic. [Online]. Available: http://www.l1-magic.org Oct 2010
[9] John G Proakis and Dimitris G Manolakis, "*Digital Signal Processing*,",ed., Pearson, 2007
[10] R. C. Gonzalez and P. Wintz, "*Digital Image Processing*," ed., Prentice Hall., 2008.
[11] T. V. Sreenivas and W. B. Kleijn, "Compressive sensingfor sparsely excited speech signals," in Proc. IEEEInt. Conf. Acoust., Speech, Signal Processing, 2009, pp.4125–4128.
[12] D. Giacobello, M. G. Christensen, J. Dahl, S. H. Jensen,and M. Moonen, "Sparse linear predictors for speechprocessing," in Proc. Interspeech, 2008.

ISBN: 978-988-19251-4-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)