AUDIO COMPRESSION USING DCT & CS

1MR. SUSHILKUMAR BAPUSAHEB SHINDE, 2PROF. MR. RAKESH MANDLIYA

1 M. Tech. (VLSI), BMCT Indore, Madhya Pradesh, India, sushilkumarshinde69@gmail.com
2 Head of EC Department, BMCT Indore, Madhya Pradesh, India, hod.electronics@bmcollege.ac.in

ABSTRACT

A large amount of techniques have been proposed to identify whether a multimedia content has been illegally tampered or not. Nevertheless, very few efforts have been devoted to identifying which kind of attack has been carried out, especially due to the large data required for this task. We propose a novel hashing scheme which exploits the paradigms of compressive sensing and distributed source coding to generate a compact hash signature, and we apply it to the case of audio content protection. At the content user side, the hash is decoded using distributed source coding tools. If the tampering is sparsifiable or compressible in some orthonormal basis or redundant dictionary, it is possible to identify the time-frequency position of the attack, with a hash size as small as 200 bits/second; the bit saving obtained by introducing distributed source coding ranges between 20% to 70%. The audio content provider produces a small hash signature by computing a limited number of random projections of a perceptual, time-frequency representation of the original audio stream; the audio hash is given by the syndrome bits of an LDPC code applied to the projections. 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 Signal, DCT, Compressive Sampling, Sparsity.

1. INTRODUCTION

Compressive Sampling (CS) is a new framework for sampling and compressing of audio and speech signal. In compressive sampling Nyquist sampling model is replaced by sparse model by assuming that signal can be represented efficiently using just few significant coefficients.

The tremendous work by Candes et.al.[3] an Donoho [4] proved that, along with implying the potential of dramatic reduction of sampling rates, power consumption and computation complexity in digital data acquisition, signal can be reconstructed with smaller than Nyquist rate.

For low power and low resolution imaging devices and or when measurement is very costly, compressive sampling is traditionally used.(e.g. Terahertz application).

But there still exits a huge gap between CS theory and its application to audio signals [13][14].How to construct a sparse audio signal, especially when CS depends on two principal: sparsity (which pertains to signal of interest), and incoherence (which pertains to sensing modality), is still unknown[6]-[8].

We have used DCT for sparse representation of an audio signal. It concentrates on the transformation content in relatively few coefficients, and it achieves a good data compression which causes its popularity [9]. Thus we can obtain a compressed version of audio signal by first obtaining a sparse representation in frequency domain, and then after processing the result with CS algorithm.

1.1 COMPRESSION SAMPLING

For increasing amount of data in our modern technology, most of data we can throw away without any perceptual loss e.g. lossy compression formats for sound, image etc. Hence question arises that why to acquire all data when most of data we will throw away? Can we directly measure only that data which is necessary? A theory of signal recovery from highly incomplete information is developed in recent series of paper [3]-[8]. Overview of results state that sparse vector x∈RN (e.g. Digital Signal) can be recovered from small number of linear measurement b=Ax∈RN or b=Ax0+e, where A is n x m matrix with far fewer rows than column (n<m) and e is measurement noise by solving a convex program.

Consider real valued signal x of length N and suppose that the basis function ψ provides k as sparse representation of x. In terms of matrix notation, we have x=ψf. In which f is sparse vector with only K non-zero elements, which can be well approximated using only k<<N non zero entries and ψ is called as sparse orthogonal basis matrix i.e. {ψ1, ψ2,...ψn} [4]

The CS theory sates that by taking only M=O(klogN) linear, non adaptive measurements shown below we can reconstruct signal x. [1],[2]:

\[ f = \Phi x = \Phi \Psi f \] (1)
Where \( Y \) represents \( M \times 1 \) sampled vector and \( \Phi \) is an \( M \times N \) measurement matrix that is incoherent with \( \Psi \), i.e., the maximum magnitude of the element in \( \Phi, \Psi \) is small [7].

Along with this information we decide to recover the signal by \( L_1 \)-minimisation is probably exact [1].

1.2 One Dimensional DCT

The most common DCT definition of a 1-D sequence of length \( N \) is

\[
C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos\left(\frac{\pi}{N} (2x + 1)u\right)
\]

for \( u = 0, 1, 2, \ldots, N - 1 \). Similarly, the inverse transformation is defined as

\[
f(x) = \frac{1}{N} \sum_{u=0}^{N-1} \alpha(u) C(u) \cos\left(\frac{\pi}{N} (2x + 1)u\right)
\]

For \( x = 0, 1, 2, \ldots, N - 1 \). In both equations (ii) and (iii) \( \alpha(u) \) is defined as

\[
\alpha(u) = \begin{cases} 
1 & \text{for } u = 0 \\
\frac{1}{\sqrt{N}} & \text{for } u \neq 0
\end{cases}
\]

It is clear from (1) that for \( u = 0 \),

\[
c(u = 0) = \frac{1}{N} \sum_{x=0}^{N-1} f(x)
\]

Thus, the first transform coefficient is the average value of the sample sequence. This value is referred to as the DC Coefficient and all other transform coefficients are called the AC Coefficients in Literature. [9]

1.3 THE TIME-FREQUENCY FILTER BANK

The MP3 standard [4] recommends the use of a high pass filter. A high pass filter allows frequencies above a given cutoff frequency to pass and does not allow lower ones to pass. In other words, it attenuates the lower frequencies. The cutoff frequency should be in the range of 2 Hz to 10 Hz.

1.4 THE POLY PHASE FILTER

The polyphase filter used in MP3 [8] is adapted from an earlier audio coder named Masking Pattern Adapted Universal Subband Integrated Coding and Multiplexing (MUSICAM). It is a cosine modulated lowpass prototype filter with uniform bandwidth parallel \( M \)-channel bandpass filter. This achieves nearly perfect reconstruction and has been called a pseudo QMF (Quadrature Mirror Filter).

2. PROPERTIES OF DCT

2.1 ENERGY COMPACITION

For highly correlated signals DCT exhibits excellent energy compaction. The uncorrelated signal has its energy spread out, whereas the energy of the correlated signal is packed into the low frequency region. Using the ability to pack input data efficiency of transformation scheme can be directly gauged into as few coefficients as possible. Because of this quantizer allows to discard coefficient with relatively small amplitudes without introducing visual distortion in reconstructed signal.

2.2 ORTHOGONALITY

IDCT basis functions are orthogonal. Thus, the inverse transformation matrix of \( A \) is equal to its transpose i.e. \( \text{inv}A=A' \). Where \( A \) is any random \( n \times n \) matrix. Therefore in addition to its decorrelation characteristics, this property results reduction in pre-computation complexity.

2.3 SYMMETRY

This is extremely useful property since it implies that the transformation matrix can be precomputed offline and applied to the signal thereby providing orders of magnitude improvement in computation efficiency.

2.4 DECORRELATION

The principle advantage of signal transformation is the removal of redundancy between neighboring pixels. This leads to uncorrelated the transform coefficients which can be encoded independently. The amplitude of autocorrelation after the DCT operation is very small hence it can be inferred that DCT exhibits excellent decorrelation properties.

2.5 SEPARABILITY

Perform DCT operation in any of the direction first and then apply on opposite direction, then also coefficient will not change.

3. METHODOLOGY

This section includes proposed techniques applied to an audio signal and described the technique for representing it the form of sparse.
speech signals which are only approximately sparse. To obtain an accurate reconstruction of such signal from highly under sampled measurement is the main issue. Ideally we have to measure all the N coefficients of f, but CS framework will allow observing a subset of these only and collecting the data.

As seen in Figure 2, the audio signal (funky.wav) is considered here for the operation in Time domain but not sparse, hence we have applied Fast Fourier Transform (FFT) which represents our signal in frequency domain and in the form of Sparse Signal as shown in Fig. 3.

Due to Matrix transformation on compressive sampling program, as describe in [10], phase angle changes because of representation in real and complex parts. Hence just applying Inverse Fourier Transform, original signal won’t be recovered.

4. RESULTS
As per above discussion, here we have taken a sample audio signal as shown in Figure 4. A spectrogram is visual representation of spectrum of frequency in sound.[15]. For better visibility and understanding of this signal, we have constructed spectrogram as shown in Figure 5. Spectrogram is nothing but graph of Time versus Normalized frequency.

As our requirement is first to generate sparse signal, we have taken DCT of audio signal shown in Figure 6. For better performance and good compression of given audio signal, we can omit the unnecessary noisy samples by thresholding. We can decide the range for thresholding as per our need. Here we have taken the range of thresholding as -0.06 to 0.04. This range of thresholding has been decided by trial and error method and selecting the threshold range which gives better output as shown in Figure 7.

Now for generation of observation vector (Figure 9), we have taken random samples from original audio signal and then reconstructed a matrix called ‘Random measurement matrix’ as shown in Figure 8. By multiplying threshold signal with random measurement matrix we get observation vector which is used for further process of reconstruction.

$L_1$ minimization is theory of signal reconstruction from highly incomplete information[16]. So authors have used $L_1$ minimization for reconstruction of audio signal. The reconstructed audio signal using $L_1$ minimization is shown in Fig. 10. Now to reconstruct this signal into original audio signal we have taken IDCT of it as shown in Figure 11 and its spectrogram is shown in Fig. 12.

In this expt., we have to consider number of samples and compression ratio is given by

$$\text{Similarity} = \frac{e^2}{E^2} \quad (iv)$$

Where $e^2$ is Matrix error given by norm of $||x-x_{rec}||$ divided by his length. And $E^2$ is matrix power given by by norm of $||x||$ divided by his length.

In Experiment-I, Sparsity is kept constant at value of 1000 and Compression Factor is kept constant at value of 0.05. By varying Block Size, Compression Ratio, SNR, PSNR is measured.

From the table 1 we can see that, for the less block size measure of compression ratio is more, but SNR and PSNR is less. But as Block size goes on increasing, measure of compression ratio goes on decreasing where as SNR and PSNR goes on increasing. For Block size 8, Compression ratio is 0.6877, SNR is 3.3517 dB and PSNR is 15.5773 dB where as for block size 512, compression ratio is 0.59688 which is low as compare to compression ratio of block size 8, SNR is 4.0212 dB and PSNR is 16.4728 dB which is high as compare to compression ratio of block size 8.

Table -1: Measure of Compression Ratio, SNR and PSNR for various values of Block size

<table>
<thead>
<tr>
<th>SR. No.</th>
<th>Block Size</th>
<th>Measure of Compression Ratio</th>
<th>SNR</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>0.6877</td>
<td>3.3517</td>
<td>15.5773</td>
</tr>
<tr>
<td>2</td>
<td>16</td>
<td>0.64371</td>
<td>3.1055</td>
<td>15.4329</td>
</tr>
</tbody>
</table>
In Experiment-II, Block Size is kept constant at value of 128 and Compression Factor is kept constant at value of 0.05. By varying Sparsity, Compression Ratio, SNR, PSNR is measured.

From the table II we can see that, for the less value of Sparsity, SNR and PSNR is less. But as Sparsity goes on increasing, SNR and PSNR goes on increasing.

For Sparsity 128, SNR is -0.5677 dB and PSNR is 11.6113 dB where as for Sparsity 2500, SNR is 20.1456 dB and PSNR is 32.3249 dB which is too high as compare to SNR and PSNR at Sparsity 128.

Table -2: Measure of Compression Ratio, SNR and PSNR for various values of Sparsity

<table>
<thead>
<tr>
<th>SR. No.</th>
<th>Sparsity</th>
<th>SNR</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>128</td>
<td>-0.5677</td>
<td>11.6113</td>
</tr>
<tr>
<td>2</td>
<td>300</td>
<td>-0.07818</td>
<td>12.1008</td>
</tr>
<tr>
<td>3</td>
<td>600</td>
<td>1.6737</td>
<td>13.8527</td>
</tr>
<tr>
<td>4</td>
<td>800</td>
<td>2.3328</td>
<td>14.5118</td>
</tr>
<tr>
<td>5</td>
<td>1000</td>
<td>3.2318</td>
<td>15.4109</td>
</tr>
<tr>
<td>6</td>
<td>1200</td>
<td>5.1285</td>
<td>17.3075</td>
</tr>
<tr>
<td>7</td>
<td>1500</td>
<td>6.8487</td>
<td>19.0277</td>
</tr>
<tr>
<td>8</td>
<td>1800</td>
<td>10.2104</td>
<td>22.3895</td>
</tr>
<tr>
<td>9</td>
<td>2000</td>
<td>14.0706</td>
<td>26.2496</td>
</tr>
<tr>
<td>10</td>
<td>2500</td>
<td>20.1459</td>
<td>32.3249</td>
</tr>
</tbody>
</table>

In Experiment-III, Block Size is kept constant at value of 128 and Sparsity is kept constant at value of 1000. By varying Compression factor (CF), Measure of Compression Ratio, SNR, PSNR is measured.

From the table III we can see that, for the less Compression factor, measure of compression ratio is more but SNR and PSNR is Compression factor 0.01. Measure of compression ratio is 0.86133, SNR is 3.395 dB and PSNR is 15.7545 dB where as for Compression factor 2.5, measure of compression ratio is 0.0019531 which is too low as compare to measure of compression ratio of Compression factor 0.01. SNR is 17.927 dB and PSNR is 32.8129 dB which is also too high.

Table -3: Measure of Compression Ratio, SNR and PSNR for various values of Compression factor (CF).

<table>
<thead>
<tr>
<th>SR. No.</th>
<th>CF</th>
<th>Measure of Compression Ratio</th>
<th>SNR</th>
<th>PSNR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.86133</td>
<td>3.395</td>
<td>15.7545</td>
</tr>
<tr>
<td>2</td>
<td>0.03</td>
<td>0.68516</td>
<td>3.9678</td>
<td>16.3307</td>
</tr>
<tr>
<td>3</td>
<td>0.05</td>
<td>0.55898</td>
<td>4.0602</td>
<td>16.2393</td>
</tr>
<tr>
<td>4</td>
<td>0.08</td>
<td>0.4375</td>
<td>3.7993</td>
<td>16.052</td>
</tr>
<tr>
<td>5</td>
<td>0.1</td>
<td>0.38125</td>
<td>3.7526</td>
<td>15.9674</td>
</tr>
<tr>
<td>6</td>
<td>0.5</td>
<td>0.080469</td>
<td>3.6443</td>
<td>16.7103</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>0.028516</td>
<td>3.9149</td>
<td>17.0527</td>
</tr>
<tr>
<td>8</td>
<td>1.5</td>
<td>0.0125</td>
<td>6.6961</td>
<td>19.4112</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>0.0054687</td>
<td>10.7931</td>
<td>24.5405</td>
</tr>
<tr>
<td>10</td>
<td>2.5</td>
<td>0.0019531</td>
<td>17.9247</td>
<td>32.8129</td>
</tr>
</tbody>
</table>

5. CONCLUSION

Input data is packed into few coefficients in DCT speech signal representation. This helps quantizer to remove coefficients with smaller amplitudes without generating audio distortion in reconstructed signal.

Compressive sampling is mainly used for compression of images but we can achieve good results by preprocessing the audio signal.

This technique can achieve a significant reduction in number of samples required to represent certain audio signal and it reduces required number of bytes for encoding.

Further improvements are possible with advanced coding techniques like Wavelet or DWT [23].

ACKNOWLEDGEMENTS

I want to give my whole sincere to my supervisor and grateful appreciation to Prof. Mr. Rakesh Mandliya, he tried her best to help me. Without her guidance I cannot bring the theories into practice. On the other hand, I want to thank all my family members and friends for their always support and spiritual motivation.

Thanks a lot!

REFERENCES


BIographies

Mr. Sushilkumar Bapusaheb Shinde completed B.E. in Electronics & Telecommunication in 2012 from SRESPC, Kopargaon, Pune University, Pune and currently pursuing M. Tech. in VLSI from BMCT, Indore, RGPV, MP.

Prof Mr. Rakesh Mandliya is Head of EC Department, BMCT Indore, Madhya Pradesh, India.