Mr.Abhang S.B, Prof.Gundal S.S / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 3, Issue 4, Jul-Aug 2013, pp.2395-2398

Sparse Audio Tampering Based On Compressive Sensing Techniques

Mr.Abhang S.B[1], Prof.Gundal S.S.[2]
[1]ME Electronics Engineering, [2]Asst.Prof. Electronics Engineering
Amrutvahini College of Engineering Sangamner (MS), 422605

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: *digital audio data authentication, robust audio fingerprinting system, Multimedia Signal Processing*

I. INTRODUCTION

With the increasing diffusion of digital multimedia contents in the last years, the possibility of tampering with multimedia contents—an ability traditionally reserved, in the case of analog signals, to few people due to the prohibitive cost of the professional equipment—has become quite a widespread practice. In addition to the ease of such manipulations, the problem of the diffusion of unauthorized copies of multimedia contents is exacerbated by security vulnerabilities and peer-to-peer sharing over the Internet, where digital contents are typically distributed and posted. This is particularly true for the case of audio files, which represent the most common example of digitally distributed multimedia contents. Some versions of the same audio piece may differ from the original because of processing, due for example to compression, resampling, or transcoding at intermediate nodes. In other cases, however, malicious attacks may occur by tampering with part of the audio stream and possibly affecting its semantic content. Examples of this second kind of attacks are the alteration of a piece of evidence in a criminal trial, or the manipulation of public opinion through the use of false wiretapping. Often, for the sake of information integrity, not only it is useful to detect whether the audio content has been modified or not, but also to identify which kind of attack has been carried out. The reasons why it is generally preferred to identify how the content has been tampered with are twofold: on one hand, given an estimate of *where* the signal was manipulated, one can establish whether or not the audio file is still meaningful for the final user; on the other hand, in some circumstances, it may be possible to recover the original semantics of the audio file.

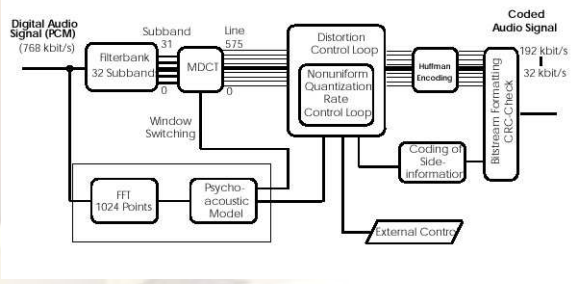


Figure 1.2: Detailed Encoder Diagram

A) 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.

B) 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 *psuedo QMF* (Quadrature Mirror Filter).

We have recently proposed a new image hashing technique [18] which exploits both the distributed source coding paradigm and the recent developments in the theory of compressive sensing. The algorithm proposed in this paper extends these ideas to the scenario of audio tampering. It also shares some similarities with the works in [15, 17]; as in [17], the hash is generated by computing random projections starting from a perceptually significant time-frequency representation of the audio signal and storing the syndrome bits obtained by Low-Density Parity-Check Codes (LDPC) encoding the quantized coefficients. With respect to [17], the proposed algorithm is novel in the following aspect: by leveraging compressive sensing principles, we are able to identify tamperings that are not sparse in the time domain only, but that can be represented by a sparse set of coefficients in some orthonormal basis or redundant dictionary. Even if the spatial models introduced in [15] could be thought of as a representation of the tampering in some dictionary, it is apparent that the compressive sensing interpretation allows much more flexibility in the choice of the sparsifying basis, since it just uses off-the-shelf basis expansions (e.g., wavelet or DCT) which can be added to the system for free.

II. RELATED WORK

We review the important concepts behind compressive sensing and distributed source coding that constitute the underlying theory of the proposed tampering identification system. In spite of the relatively large amount of literature published on these fields in the past few years, this is a very concise introduction; for a more detailed and exhaustive explanation the interested reader may refer to for compressive sensing and to for distributed source coding.

2.1. Compressive Sampling

Compressive sampling (or compressed sensing) is a new paradigm which asserts that it is possible to perfectly recover a signal from a limited number of incoherent, nonadaptive linear measurements, provided that the signal admits a sparse representation in some orthonormal basis or redundant dictionary, that is, it can be represented by a small number of nonzero coefficients in some basis expansion. Let $\mathbf{x} \in \mathbb{R}^n$ be the signal to be acquired, and $\mathbf{y} \in \mathbb{R}^m$, $m < n$, a number of linear random projections (measurements) obtained as $\mathbf{y} = \mathbf{A}\mathbf{x}$. In general, given the prior knowledge that $\mathbf{x}$ is k-sparse, that is, that only k out of its n coefficients are different from zero, one can recover $\mathbf{x}$ by solving the following optimization problem:

$$\min \|\mathbf{x}\|_0 \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

where $\|\cdot\|_0$ simply counts the number of nonzero elements of $\mathbf{x}$. This program can correctly recover a k-sparse signal from $m = k + 1$ random samples. Unfortunately, such a problem is NP hard, and it is also difficult to solve in practice for problems of moderate size.

2.2. Distributed Source Coding

Consider the problem of communicating a continuous random variable X. Let Y denote another continuous random variable correlated to X. In a distributed source coding setting, the problem is to decode X to its quantized reconstruction $\hat{X}$ given a constraint on the distortion measure $D = E[d(X, \hat{X})]$ when the side information Y is available only at the decoder. Let us denote by $R_{X|Y}(D)$ the rate-distortion function for the case when Y is also available at the encoder, and by $R_{X|Y}^{WZ}(D)$ the case when only the decoder has access to Y. The Wyner-Ziv theorem states that, in general, $R_{X|Y}^{WZ}(D) \geq R_{X|Y}(D)$ but $R_{X|Y}^{WZ}(D) = R_{X|Y}(D)$ for Gaussian memoryless sources and mean square error (MSE) as distortion measure.

The Wyner-Ziv theorem has been applied especially in the area of video coding under the name of distributed video coding (DVC), where the source X (pixel values or DCT coefficients) is quantized with 2^J levels, and the J bitplanes are independently encoded, computing parity bits by means of a turbo encoder. At the decoder, parity bits are used together with the side information Y to "correct" Y into a quantized version $\hat{X}$ of X, performing turbo decoding, typically starting from the most significant bitplanes. To this end, the decoder needs to know the joint probability density function (pdf) $p_{XY}(X, Y)$. More recently, LDPC codes have been adopted instead of turbo codes Although the rate-distortion performance of a practical DSC codec strongly depends on the actual implementation employed, it is yet possible to approximately quantify the gain obtained by introducing a Wyner-Ziv coding paradigm, in order to estimate the bit saving produced in the hash signature. Let X and Y be zero mean, i.i.d. Gaussian variables with variance, respectively, σ_X^2 and σ_Y^2; also, let σ_N^2 be the variance of the innovation noise $N = Y - X$. Classical information theory asserts that the rate expressed in bits per sample for a given distortion level D, in the case of a Gaussian source X is given by

$$R_X(D) = \frac{1}{2}\log_2\frac{\sigma_X^2}{D}.$$

σ_X^2 relates to the energy of the original signal, while σ_N^2 to the energy of the tampering. Equation (8) shows that the advantage of using a DSC approach with respect to a traditional quantization and encoding becomes consistent when the signal and the side information are well correlated, that is, when the energy of the tampering is small relative to the energy of the original sound.

III. TAMPERING MODEL

Before describing in more detail the architecture of the system, we need to set up a model for sparse tampering. Let $\mathbf{x} \in \mathbb{R}^n$ be the original signal; we model the effect of a sparse tampering $\mathbf{e} \in \mathbb{R}^n$ as

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{e},$$

where $\tilde{\mathbf{x}}$ is the modified signal received by the user. We postulate without loss of generality that $\mathbf{e}$ has only k nonzero components (in fact, it suffices for $\mathbf{e}$ to be sparse or compressible in some basis or frame). Let $\mathbf{y} = \mathbf{A}\mathbf{x}$ be the random measurements of the original signal and $\tilde{\mathbf{y}} = \mathbf{A}\tilde{\mathbf{x}}$ be the projections of the tampered signal; clearly, the relation between the tampering and the measurements is given by

$$\mathbf{b} = \tilde{\mathbf{y}} - \mathbf{y} = \mathbf{A}(\tilde{\mathbf{x}} - \mathbf{x}) = \mathbf{A}\mathbf{e}.$$

If the sensing matrix $\mathbf{A}$ is chosen such that it satisfies the RIP, we have that

$$\|\mathbf{b}\|_2 = \|\mathbf{A}\mathbf{e}\|_2 \approx \sqrt{\frac{m}{n}}\,\|\mathbf{e}\|_2,$$

and thus we are able to approximate the energy of the tampering from the projections computed at the decoder and the encoder-side projections reconstructed exploiting the hash. This fact comes out to be very useful to estimate the energy of the tampering at the CU side and will be exploited in Section 4. Furthermore in order to apply the Wyner-Ziv theorem, we need $\mathbf{b}$ to be i.i.d. Gaussian with zero mean. This has been verified through experimental simulations on several tampering examples. Indeed, a theoretical justification can be provided by invoking the central limit theorem, since each element $b_i = \sum_{j=1}^{n} A_{ij}e_j$ is the sum of random variables whose statistics are not explicitly modeled.

3.1 Hash Decoding and Tampering Identification

The CU receives the (possibly tampered) audio stream $\tilde{\mathbf{X}}$ and requests the syndrome bits and the random seed of the hash $\mathcal{H}(\mathbf{X}, S)$ from the authentication server. On each user's request, a different seed S is used in order to avoid that a malicious attack could exploit the knowledge of the nullspace of $\mathbf{A}$.

(1) Frame-Based Subband Log-Energy Extraction

A perceptual, time-frequency representation of the signal $\tilde{\mathbf{X}}$ received by the CU is computed using the same algorithm described above for the CP side. At this step, the vector $\tilde{\mathbf{x}}$ is produced.

(2) Random Projections

A set of m linear random measurements $\tilde{\mathbf{y}} = \mathbf{A}\tilde{\mathbf{x}}$ are computed using a pseudorandom matrix $\mathbf{A}$ whose entries are drawn from a Gaussian distribution with the same seed S as the encoder.

(3) Wyner-Ziv Decoding

A quantized version $\tilde{\mathbf{y}}$ is obtained using the hash syndrome bits and $\tilde{\mathbf{y}}$ as side information. LDPC decoding is performed starting from the most significant bitplane.

IV. EXPERIMENTAL RESULTS

We have carried out some experiments on 32 seconds of speech audio data, sampled at 44100 Hz and 16 bits per sample. The test audio consists of a piece of a newspaper article read by a speaker; the recording is clean but for some noise added at a few time instants, including the high frequency noise of a shaken key ring, the wide-band noise of some crumpling paper, and some impulsive noise in the form of coughs of the speaker. We have set the size of the audio frame to $F = 11025$ samples (0.25 seconds), and the number of Mel frequency bands to $U = 32$, obtaining a total of 128 audio frames corresponding to $n = 4096$ log-energy coefficients. We have then assembled a testbed considering 3 kinds of tampering.

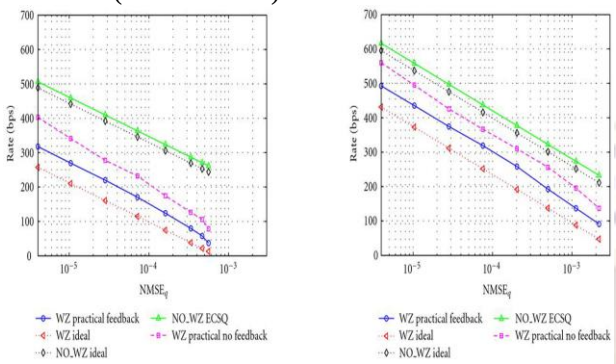
4.1 Rate-Distortion Performance Of The Hash Signature

As described in Section 4, we use distributed source coding for reducing the payload due to the hash. In this section, we want to quantify the bit-saving obtained with Wyner-Ziv coding of the hash. In order to do so, we have compared the rate distortion function of Wyner-Ziv (WZ) coding and of hash direct quantization and transmission, that is, without using DSC (NO-WZ). Figure 4depicts these two situations for the cases of the frequency and time domain tampering. In both the two graphs, the value of quantization MSE has been normalized by the energy of the measurements $\mathbf{y}$, in order to make the result comparable with other possible manipulations

$$\mathrm{NMSE}_q = \frac{\|\mathbf{y} - \hat{\mathbf{y}}\|_2^2}{\|\mathbf{y}\|_2^2}. \quad (20)$$

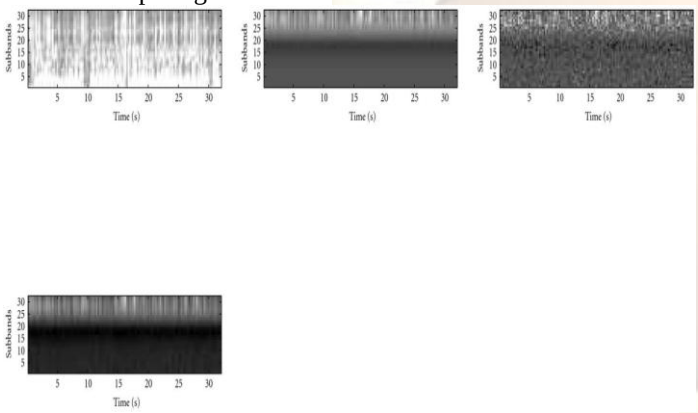
The bold-dotted lines represents the theoretical WZ rate-distortion curve of the measurements stated in (7). The bold solid and dashed lines represent instead the actual rate-distortion behavior obtained by using a practical WZ codec, either using the feedback channel or directly

estimating at the encoder side the rate as explained in Section 6. For comparison, we have also plotted the rate-distortion functions of an ideal NO-WZ uniform quantizer (Shannon's bound), drawn as a thin-dotted line, and the rate-distortion curve of an entropy-constrained scalar quantization (ECSQ), which is a well studied and effective practical quantization scheme (thin-solid line).



4.2 Choice of the Best Tampering Reconstruction

In practice, the tampering may be sparse or compressible in more than one basis: this may be the case, for instance, of piece-wise polynomials signals which are generally sparse in several wavelet expansions. When this situation occurs, multiple tampering reconstructions are possible, and at the CU side there is an ambiguity about what is the best tampering estimation.



To have a visual insight of the effect of different bases in the tampering reconstruction, we have drawn in Figure 5 the log-energy spectrum of the original audio signal and of the frequency-localized (F) tampering, followed by the log-energy spectrum of the tampering reconstructed in two different domains using a hash rate of 200 bps. It apparent from the figure that the quality of the estimated tampering reconstructed using 2D-DCT considerably overcomes the one obtained in the log-energy domain.

V. CONCLUSION

We presented a hash-based tampering identification system for detecting and identifying illegitimate manipulations in audio files. The algorithm works with sparse modifications, leveraging the recent compressive sensing results for reconstructing the tampering from a set of random nonadaptive measurements. Perhaps the most distinctive feature of the proposed system is its ability to reconstruct a tampering that is sparse in some orthonormal basis or frame, without knowing at the CP side the actual content alteration. 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.

REFERENCES

[1] M Steinebach, J Dittmann, Watermarking-based digital audio data authentication. EURASIP Journal on Applied Signal Processing **2003**(10), 1001–1015 (2003). Publisher Full Text

[2] J Fridrich, Image watermarking for tamper detection. Proceedings of IEEE International Conference on Image Processing (ICIP '98), October 1998, Chicago, Ill, USA **2**, 404–408

[3] JJ Eggers, B Girod, Blind watermarking applied to image authentication. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01), May 2001, Salt Lake, Utah, USA **3**, 1977–1980

[4] D Kundur, D Hatzinakos, Digital watermarking for telltale tamper proofing and authentication. Proceedings of the IEEE **87**(7), 1167–1180 ,Y-C Lin, D Varodayan, B Girod, Image authentication based on distributed source coding. Proceedings of IEEE