Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2009, Article ID 158982, 12 pages
doi:10.1155/2009/158982

Research Article

Identification of Sparse Audio Tampering Using Distributed Source Coding and Compressive Sensing Techniques

G. Valenzise, G. Prandi, M. Tagliasacchi, and A. Sarti

Dipartimento di Elettronica e Informazione, Politecnico di Milano, P.zza Leonardo da Vinci, 32 20133 Milano, Italy

Correspondence should be addressed to G. Valenzise, valenzise@elet.polimi.it

Received 16 May 2008; Revised 30 September 2008; Accepted 20 November 2008

Recommended by Anthony Vetro

The increasing development of peer-to-peer networks for delivering and sharing multimedia files poses the problem of how to protect these contents from unauthorized manipulations. In the past few years, 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 apply it to the case of audio content protection. 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. 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%.

1. 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.

In the past literature, the aim of distinguishing legitimately modified copies from manipulations of a multimedia file has been addressed with two kinds of approaches: watermarks and media hashes. Both approaches have been extensively applied to the case of image content types, while fewer systems have been proposed for the case of audio signals. Digital watermarking techniques embed information directly into the media data to ensure both data integrity and authentication. Even if digital watermarks can be categorized based on several properties, such as robustness, security,

complexity, and invertibility [1], a common taxonomy is to distinguish between *robust* and *fragile* watermarks. It is the latter category that is particularly useful for checking the integrity of an audio file; a fragile watermark is a mark that is easily altered or destroyed when the host data is modified through some transformation, either legitimate or not. If the watermark is designed to be robust with respect to legitimate, perceptually irrelevant modifications (e.g., compression or resampling), and at the same time to be fragile with respect to perceptually and semantic significant alterations, then it is a *content-fragile* watermark [1]. With this scheme, a possible tampering can be detected and localized by identifying the damage to the extracted watermark. Examples of this approach for the case of image content types are given in [2, 3]. The authors of [4] propose an image authentication scheme that is able to localize tampering, by embedding a watermark in the wavelet coefficients of an image. If a tampering occurs, the system provides information on specific frequencies and space regions of the image that have been modified. This allows the user to make application-dependent decisions concerning whether an image, which is JPEG compressed for instance, still has credibility. A similar idea, also working on the signal wavelet domain, has been applied to audio in [5], with the aim of copyright verification and tampering identification. The image watermarking system devised in [6] inserts a fragile watermark in the least significant bits of the image on a block-based fashion; when a portion of the image is tampered with, only the watermark in the corresponding blocks is destroyed, and the manipulation can be localized. Celik et al. [7] extend this method by inserting the watermark in a hierarchical way, to improve robustness against vector quantization attacks. In [8], image protection and tampering localization is achieved through a technique called "cocktail watermarking"; two complementary watermarks are embedded in the original image to improve the robustness of the detector response, while at the same time enabling tampering localization. The same ideas have been applied by the authors to the case of sounds [9], by inserting the watermark in the host audio FFT coefficients. For a more exhaustive review of audio watermarking for authentication and tampering identification see Steinebach and Dittmann [1].

Despite their widespread diffusion as a tool for multimedia protection, watermarking schemes suffer from a series of disadvantages: (1) watermarking authentication is not backward compatible with previously encoded contents (unmarked contents cannot be authenticated later by just retrieving the corresponding hash); (2) the original content is distorted by the watermark; (3) the bit rate required to compress a multimedia content might increase due to the embedded watermark. An alternative solution for authentication and tampering identification is the use of *multimedia hashes*. Unlike watermarks, content hashing embeds a signature of the original content as part of the header information, or can provide a hash separately from the content upon a user's request. Multimedia hashes are inspired by cryptographic digital signatures, but instead of being sensitive to single-bit changes, they are supposed to offer proof of perceptual integrity. Despite some audio hashing systems (also named *audio fingerprinting*) being proposed in the past few years [10–12], most of the previous research, as for the case of watermarking, has concentrated on the case of images [13, 14]. In [10], the authors build audio fingerprints by collecting and quantizing a number of robust and informative features from an audio file, with the purpose of audio identification as well as fast database lookup. Haitsma and Kalker [11] build audio fingerprints robust to legitimate content modifications (mp3 compression, resampling, moderate time, and pitch scaling), by dividing the audio signal in highly overlapping frames of about 0.3 seconds; for each frame, they compute a frequency representation of the signal through a filter bank with logarithmic spacing among the bands, in order to resemble the human auditory system (HAS). The redundance of musical sounds is exploited by taking the differences between subbands in the same frame, and between the same subbands in adjacent time instants; the resulting vector is quantized with one bit, and similarities between each short fingerprint are computed through the Hamming distance. By concatenating all the fingerprints of each frame, a global hash is obtained, which is used next to efficiently query a song database of previously encoded fingerprints. Though in principle such an approach could be used for identifying possible localized tampering in the audio stream, the authors do not explicitly address this problem. An excellent review of algorithms and applications of audio fingerprinting is presented in [12].

To the best of the authors' knowledge, no audio hashing technique has been used up to now with the purpose of detecting and localizing unauthorized audio tampering. One of the main reasons of that is probably the great amount of bits of the audio hashes required for enabling the identification of the tampering, when traditional fingerprinting approaches as the ones described above are employed. In fact, in order to limit the rate overhead, the size of the hash needs to be as small as possible. At the same time, the goal of tampering localization calls for increasing the hash size, in order to capture as much as possible about the original multimedia object. Recently, Lin et al. have proposed a new hashing technique for authentication [14] and tampering localization [15] for images, which produce very short hashes by leveraging distributed source coding theory. In this system, the hash is composed of the Slepian-Wolf encoding bitstream of a number of quantized random projections of the original image; the content user (CU) computes its own random projections on the received (and possibly tampered) image, and uses them as a side information to decode the received hash. By setting some maximum predefined tampering level on the received image (e.g., a minimum tolerated PSNR between the original and the forged image is allowed), it is possible to transmit the hash without the need of a feedback channel, performing rate allocation at the encoder side (a similar bit allocation technique has been adopted by the authors also in the context of reduced-reference image quality assessment [16]). When decoding succeeds, it is possible to identify tampered regions of the image, at the cost of additional hash bits. This scheme has been applied also to the case of audio files [17]; instead

of random projections of pixels, the authors compute for each signal frame a weighted spectral flatness measure, with randomly chosen weights, and encode this information to obtain the hash. Though this scheme applies well to the authentication task (which can be attained with a hash overhead less than 100 bits/second), it is not clear how to extend the application to identification of general kinds of tampering.

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.

To clear up which are the capabilities and the limitations of the proposed system, Figure 1 shows an example of malicious tampering with an audio signal. This demonstration has been carried out on a piece of audio speech, with a length of approximately 2 seconds, read from a newspaper by a speaker. The whole recording, which is about 32 seconds long, has also been used as a proof of concept to present some experimental results on the system in Section 7. Figure 1(a) shows the original waveform, which corresponds to the Italian sentence "un sequestro da tredici milioni di euro" (a confiscation of thirteen million euros). This sentence has been tampered with in order to substitute the words "tredici milioni" (thirteen million) with "quindici miliardi" (fifteen billion), see Figure 1(b). In order to compute the hash, as explained in Section 4, we compute a coarse-scale perceptual time-frequency map of the signal (in this case, with a temporal resolution of 1/4 seconds). From the received tampered waveform and from the information of the hash, the user is able to identify the tampering (Figure 1(d)).

The rest of the paper is organized as follows: Section 2 provides the necessary background information about compressive sensing and distributed source coding; Section 3 describes the tampering model; Section 4 gives a detailed description of the system; Section 6 describes how it is possible to estimate the rate of the hash at the encoder without feedback channel or training; the tampering identification algorithm is tested against various kinds of attacks in Section 7, where also the different bit-rate requirements for the hash with or without distributed source coding are compared; finally, Section 8 draws some concluding remarks.

2. Background

In this section, 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 [19–21] for compressive sensing and to [22–24] for distributed source coding.

2.1. Compressive Sampling (CS). 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:

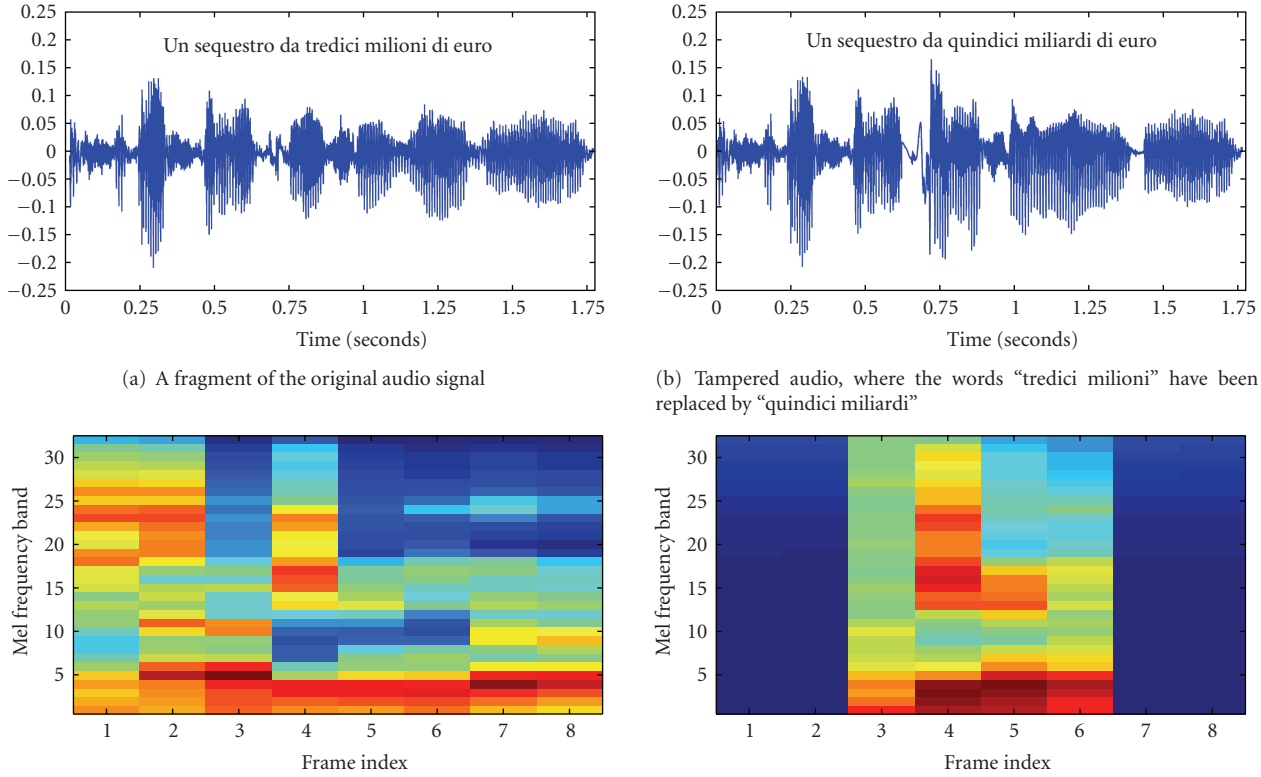
$$\min \|\mathbf{x}\|_0 \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}, \tag{1}$$

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 [25]. Unfortunately, such a problem is NP hard, and it is also difficult to solve in practice for problems of moderate size.

To overcome this exhaustive search, the compressive sampling paradigm uses special measurement matrices $\mathbf{A}$ that satisfy the so-called *restricted isometry property* (RIP) of order k [21], which says that all subsets of k columns taken from $\mathbf{A}$ are in fact nearly orthogonal or, equivalently, that linear measurements taken with $\mathbf{A}$ approximatively preserve the Euclidean length of k-sparse signals. This in turn implies that k-sparse vectors cannot be in the null space of $\mathbf{A}$, a fact that is extremely useful, as otherwise there would be no hope of reconstructing these vectors. Merely verifying that a given $\mathbf{A}$ has the RIP according to the definition is combinatorially complex; however, there are well-known cases of matrices that satisfy the RIP, obtained for instance by sampling i.i.d. entries from the normal distribution with mean 0 and variance $1/n$. When the RIP holds, then the following linear program gives an accurate reconstruction

$$\min \|\mathbf{x}\|_1 \quad \text{s.t. } \mathbf{y} = \mathbf{A}\mathbf{x}. \tag{2}$$

The solution of (2) is the same as the one of (1) provided that the number of measurements satisfy $m \geq C \cdot k \log_2(n/k)$, where C is some small positive constant. Moreover, if $\mathbf{x}$ is not exactly sparse, but it is at least *compressible* (i.e., its coefficients decay as a power law), then solving (2) guarantees



(a) A fragment of the original audio signal

(b) Tampered audio, where the words "tredici milioni" have been replaced by "quindici miliardi"

(c) A coarse-scale perceptual time-frequency map of the original signal, from which the hash signature is computed

(d) The tampering in the perceptual time-frequency domain as estimated by the proposed algorithm

Figure 1: An example of the result of the proposed audio tampering identification, applied to a fragment of speech read from a newspaper.

that the quality of the recovered signal is as good as if one knew ahead of time the location of the k largest values of $\mathbf{x}$ and decided to measure those directly [21]. These results also hold when the signal is not sparse as is, but it has a sparse representation in some orthonormal basis. Let $\boldsymbol{\Psi} \in \mathbb{R}^{n\times n}$ denote an orthonormal matrix, whose columns are the basis vectors. Let us assume that we can write $\mathbf{x} = \boldsymbol{\Psi}\boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a k-sparse vector. Clearly, (2) is a special case of this instance, when $\boldsymbol{\Psi}$ is the identity matrix. Given the measurements $\mathbf{y} = \mathbf{A}\mathbf{x}$, the signal $\mathbf{x}$ can be reconstructed by solving the following problem:

$$\min \|\boldsymbol{\alpha}\|_1 \quad \text{s.t. } \mathbf{y} = \mathbf{A}\boldsymbol{\Psi}\boldsymbol{\alpha}. \tag{3}$$

Problem (3) can be solved without prior knowledge of the actual sparsifying basis $\boldsymbol{\Psi}$ for different test bases, until a sparse reconstruction $\boldsymbol{\alpha}$ is obtained.

In most practical applications, measurements are affected by noise (e.g., quantization noise). Let us consider noisy measurements $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$, where $\mathbf{z}$ is a norm-bounded noise, that is, $\|\mathbf{z}\|_2 \leq \epsilon$. An approximation of the original signal $\mathbf{x}$ can be obtained by solving the modified problem:

$$\min \|\boldsymbol{\alpha}\|_1 \quad \text{s.t. } \|\mathbf{y} - \mathbf{A}\boldsymbol{\Psi}\boldsymbol{\alpha}\|_2 \leq \epsilon. \tag{4}$$

Problem (4) is an instance of a second-order cone program (SOCP) [26] and can be solved in $O(n^3)$ time. Several fast algorithms have been proposed in the literature that attempt to find a solution to (4). In this work, we adopt the SPGL1 algorithm [27], which is specifically designed for large-scale sparse reconstruction problems.

2.2. Distributed Source Coding (DSC). 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 $R^{\text{WZ}}_{X|Y}(D)$ the case when only the decoder has access to Y. The Wyner-Ziv theorem [23] states that, in general, $R^{\text{WZ}}_{X|Y}(D) \geq R_{X|Y}(D)$ but $R^{\text{WZ}}_{X|Y}(D) = R_{X|Y}(D)$ 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 [28, 29].

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 [30] 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

$$R_X(D) = \frac{1}{2}\log_2 \frac{\sigma_X^2}{D}. \tag{5}$$

The rate-distortion function for the case of Wyner-Ziv encoding, when the conditions of the theorem are satisfied, is

$$R_{X|Y}^{\mathrm{WZ}}(D) = \frac{1}{2}\log_2 \frac{\sigma_X^2 \sigma_N^2}{D(\sigma_X^2 + \sigma_N^2)} \tag{6}$$

which becomes, in the hypothesis that $\sigma_X^2 \gg \sigma_N^2$, approximatively equal to the rate needed to encode the innovation N

$$R_{X|Y}^{\mathrm{WZ}}(D) \approx \frac{1}{2}\log_2 \frac{\sigma_N^2}{D}. \tag{7}$$

Subtracting (7) from (5), we obtain the expected coding gain due to Wyner-Ziv coding

$$\Delta R_{\mathrm{WZ}} = \frac{1}{2}\log_2 \frac{\sigma_X^2}{\sigma_N^2}. \tag{8}$$

As we will see in Section 4, 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.

3. 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

$$\widetilde{\mathbf{x}} = \mathbf{x} + \mathbf{e}, \tag{9}$$

where $\widetilde{\mathbf{x}}$ 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 $\widetilde{\mathbf{y}} = \mathbf{A}\widetilde{\mathbf{x}}$ be the projections of the tampered signal; clearly, the relation between the tampering and the measurements is given by

$$\mathbf{b} = \widetilde{\mathbf{y}} - \mathbf{y} = \mathbf{A}(\widetilde{\mathbf{x}} - \mathbf{x}) = \mathbf{A}\mathbf{e}. \tag{10}$$

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

$$\|\mathbf{b}\|_2 = \|\mathbf{A}\mathbf{e}\|_2 \approx \sqrt{\frac{m}{n}}\|\mathbf{e}\|_2, \tag{11}$$

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.

4. Description of the System

The proposed tampering detection and localization scheme is depicted in Figure 2. The general architecture of the system is composed by two actors: on one hand, there is the *content producer* (CP), which is the entity that publishes or distributes the legitimate and authentic copies of the original audio content. On the other hand, there is the CU, which is the consumer of the audio content released by the CP. The CP disseminates copies of the original content $\mathbf{X} \in \mathbb{R}^N$, where N is the total number of audio samples of the signal, through possibly untrusted intermediaries, which may tamper with the authentic file manipulating its semantics; at the same time, the CU may get its own copy $\widetilde{\mathbf{X}}$ of the audio file from nodes different from the starting CP. In order to protect the integrity of the multimedia content, the CP builds a small hash signature $\mathcal{H}$ of the audio signal. To perform content authentication, the user sends a request for the hash signature to an authentication server, which is supposed to be trustworthy. By exploiting the hash, the user can estimate the distortion of the received content $\widetilde{\mathbf{X}}$ with respect to the original $\mathbf{X}$. Furthermore, if the tampering is sparse in some basis expansion, the system produces a tampering estimation $\hat{\mathbf{e}}$ which identifies the attack in the time-frequency domain. In the following, we detail the hash generation procedure at the CP side and the tampering identification at the CU side.

4.1. Generation of the Hash Signature. At the CP side, given the audio stream $\mathbf{X}$ and a random seed S, the encoder generates the hash signature $\mathcal{H}(\mathbf{X}, S)$ as follows.

(1) Frame-Based Subband Log-Energy Extraction. The original single-channel audio stream $\mathbf{X}$ is partitioned into

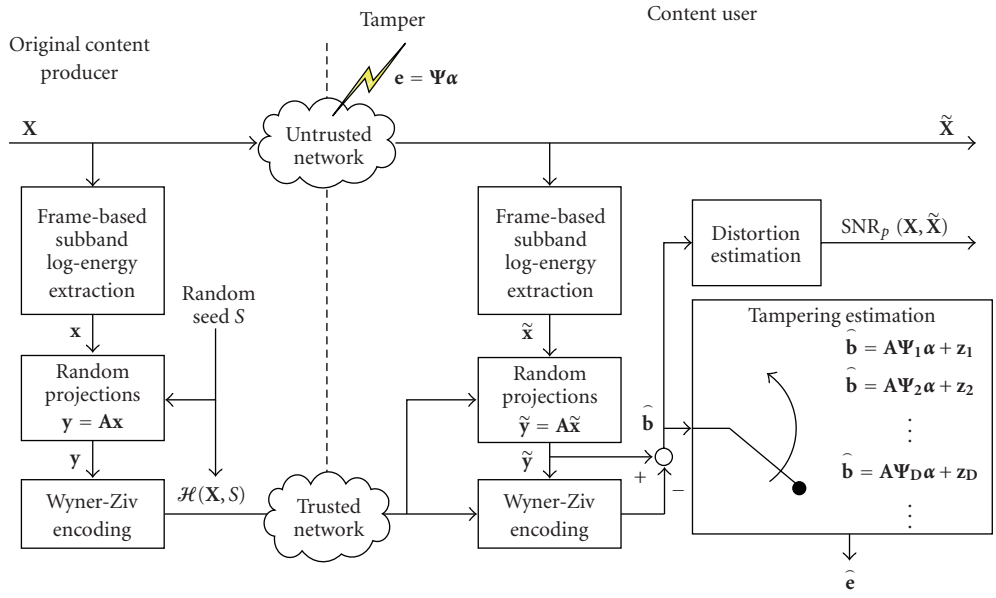


Figure 2: Block diagram of the proposed tampering identification scheme.

nonoverlapping frames of length F samples. The power spectrum of each frame is subdivided into U Mel frequency subbands [31], and for each subband the related spectral log-energy is extracted. Let $h_{f,u}$ be the energy value for the uth band at frame f. The corresponding log-energy value is computed as follows:

$$x_{f,u} = \log\left(1 + h_{f,u}\right). \tag{12}$$

The values $x_{f,u}$ provide a time-frequency perceptual map of the audio signal (see Figure 1). The log-energy values are "rasterized" as a vector $\mathbf{x} \in \mathbb{R}^n$, where $n = UN/F$ is the total number of log-energy values extracted from the audio stream.

(2) Random Projections. A number of linear random projections $\mathbf{y} \in \mathbb{R}^m$, $m < n$, is produced as $\mathbf{y} = \mathbf{A}\mathbf{x}$. The entries of the matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ are sampled from a Gaussian distribution $\mathcal{N}(0, 1/n)$, using some random seed S, which will be sent as part of the hash to the user.

(3) Wyner-Ziv Encoding. The random projections $\mathbf{y}$ are quantized with a uniform scalar quantizer with step size Δ. As mentioned in Section 1, to reduce the number of bits needed to represent the hash, we do not send directly the quantization indices. Instead, we observe that the random projections computed from the possibly tampered audio signal will be available at the decoder side. Therefore, we can perform lossy encoding with side information at the decoder, where the source to be encoded is $\mathbf{y}$ and the "noisy" random projections $\widetilde{\mathbf{y}} = \mathbf{A}\widetilde{\mathbf{x}}$ play the role of the side information. The vector $\widetilde{\mathbf{x}}$ contains the log-energy values of the audio signal received at the decoder. With respect to the distributed source coding setting illustrated in Section 2.2, we have $X = \mathbf{y}$, $Y = \widetilde{\mathbf{y}}$, $N = \mathbf{b} = \widetilde{\mathbf{y}} - \mathbf{y}$. Following the approach widely adopted in the literature on distributed video coding [24], we perform bitplane extraction on the quantization bin indices. Then each bitplane vector is LDPC coded to create the hash.

4.2. Hash Decoding and Tampering Identification. The CU receives the (possibly tampered) audio stream $\widetilde{\mathbf{X}}$ 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 $\mathbf{A}$ [14].

(1) Frame-Based Subband Log-Energy Extraction. A perceptual, time-frequency representation of the signal $\widetilde{\mathbf{X}}$ received by the CU is computed using the same algorithm described above for the CP side. At this step, the vector $\widetilde{\mathbf{x}}$ is produced.

(2) Random Projections. A set of m linear random measurements $\widetilde{\mathbf{y}} = \mathbf{A}\widetilde{\mathbf{x}}$ 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 $\widehat{\mathbf{y}}$ is obtained using the hash syndrome bits and $\widetilde{\mathbf{y}}$ as side information. LDPC decoding is performed starting from the most significant bitplane.

(i) If a feedback channel is available, decoding always succeeds, unless an upper bound is imposed on the maximum number of hash bits.

(ii) Conversely, if the actual distortion between the original and the tampered signal is higher than the maximum tolerated distortion determined by the original CP, decoding might fail.

(4) Distortion Estimation. If Wyner-Ziv decoding succeeds, an estimate of the distortion in terms of a perceptual signal-to-noise ratio is computed using the projections of the subsampled energy spectrum of the tampering. Let $\hat{\mathbf{b}} = \tilde{\mathbf{y}} - \hat{\mathbf{y}}$ be the projections of the subsampled energy spectrum of the tampering; we define the perceptual signal-to-noise ratio (SNR_P) of the received audio stream as

$$\mathrm{SNR}_P = 10 \log_{10} \frac{||\hat{\mathbf{y}}||_2^2}{||\hat{\mathbf{b}}||_2^2} \ [\mathrm{dB}]. \tag{13}$$

This definition needs some further interpretation. In fact, we compute the SNR_P from the projections in place of the whole time-frequency perceptual map of both the signal and the tampering. This is justified by the energy conservation principle stated in (11) and by the fact that, at the CU side, no information about the authentic audio content is available; hence, this is an approximation of the actual SNR_P, which uses the quantized projections obtained by decoding the hash signature, in the reasonable hypothesis that $\|\hat{\mathbf{y}}\| \approx \|\mathbf{y}\|$ and $\|\hat{\mathbf{b}}\| \approx \|\mathbf{b}\|$.

(5) Tampering Estimation. If the tampering can be represented by a sparse set of coefficients in some basis $\boldsymbol{\Psi}_i$, it can be reconstructed starting from the random projections $\mathbf{b} = \tilde{\mathbf{y}} - \hat{\mathbf{y}}$ by solving the following optimization problem, as anticipated in Section 2.1:

$$\min \|\boldsymbol{\alpha}\|_1 \quad \text{s.t.} \ \left|\left|\hat{\mathbf{b}} - \mathbf{A}\boldsymbol{\Psi}_i\boldsymbol{\alpha}\right|\right|_2 \leq \epsilon. \tag{14}$$

For a given orthonormal basis $\boldsymbol{\Psi}_i$, the expansion of the tampering in that basis, that is, $\boldsymbol{\alpha}_i = \boldsymbol{\Psi}_i^T(\mathbf{x} - \tilde{\mathbf{x}})$, might not be sparse enough with respect to the number of available random projections m and the optimization algorithm might not converge to a feasible solution. In such cases, it is not possible to perform tampering identification, and a different orthonormal basis $\boldsymbol{\Psi}_j$, $j \neq i$ is tested. If the optimization algorithm does not converge for any of the tested bases, the tampering is declared to be nonsparse. This is the case, for example, of quantization noise introduced by audio compression. If the reconstruction succeeds for more than one basis, we choose the one in which the tampering is the sparsest. While, in principle, this just means that we should take the basis that returns the smallest ℓ_0 metrics, we have in practice to cope with reconstruction noise, which in fact prevents the recovered tampering to be exactly sparse. A simple solution is to select the basis that gives the smallest ℓ_1 norm; however, this approach has the drawback of being too sensitive toward high values of the coefficients (e.g., due to different dynamic ranges in the transform domains). As experimentally shown in Section 7.2, this bias has the side-effect that selecting the minimum ℓ_1 norm reconstruction does not ensure that one is performing the best possible tampering estimation. A more effective heuristic is to use some ℓ_p metrics, with $0 < p < 1$, or similar norms, as the ones devised in [32]. In our experiments, we have computed the norm of the coefficients $\boldsymbol{\alpha}$ as

$$\|\boldsymbol{\alpha}\| = \sum_{i=1}^{m} \arctan\left(\frac{|\alpha_i|}{\delta}\right), \tag{15}$$

where δ has been set so that $\arctan(1/\delta) = 1$.

5. Choice of the Hash Parameters

In the hash construction procedure, there are two parameters that influence the quality of tampering estimation. The number of random projections m used to build the hash, and the number of bitplanes J which determines the distortion due to quantization on the reconstructed measurements at the user side. In this section we analyze the tradeoff between the rate needed to encode the hash, which also depends on the maximum allowed tampering level as explained in Section 6, and the accuracy of the tampering estimation; a larger number of bitplanes J and of measurements m correspond to a higher quality of tampering estimation, and at the same time to a higher rate spent for the hash. In order to find an optimal tradeoff between m and J, we conducted Monte carlo simulations on a generic sparse signal $\mathbf{x}$, with two different sparsity levels k/n. We evaluate the goodness of the tampering estimation by calculating the reconstruction normalized MSE (NMSE_R) between the original k-sparse signal $\mathbf{x}$ and its approximation $\hat{\mathbf{x}}$ obtained by solving problem (4)

$$\mathrm{NMSE}_R = \frac{||\hat{\mathbf{x}} - \mathbf{x}||_2^2}{||\mathbf{x}||_2^2}. \tag{16}$$

The noise $\mathbf{z} = \hat{\mathbf{x}} - \mathbf{x}$ in (4) in this case corresponds to quantization noise, which is uniformly distributed between $-\Delta/2$ and $\Delta/2$, where Δ is the quantization step size. We measure the impact of quantization noise by measuring the signal-to-quantization noise ratio

$$\mathrm{SNR}_{\mathbf{y}} = 10 \log_{10} \frac{||\mathbf{y}||_2^2}{||\mathbf{y} - \hat{\mathbf{y}}||_2^2}, \tag{17}$$

where $\hat{\mathbf{y}}$ is the quantized version of the random projections $\mathbf{y} = \mathbf{A}\mathbf{x}$. As for the reconstruction basis, $\boldsymbol{\Psi}$, we just assign $\boldsymbol{\Psi} = I$ in (4), that is, we assume that the signal is sparse as is, or equivalently that some oracle has told us the optimal sparsifying basis in advance. Figure 3 shows the NMSE_R contour set for two levels of sparsity ($k/n = 0.15$ and $k/n = 0.25$) as a function of the number of projections m and of the quantization distortion of the measurements ($\mathrm{SNR}_{\mathbf{y}}$). We observe a graceful improvement of the performance by increasing either m or $\mathrm{SNR}_{\mathbf{y}}$. For the same values of the parameters, the normalized MSE of the reconstructed signal is lower for sparser signals ($k/n = 0.15$). This is justified by the CS result on the number of projections which requires $m \geq C \cdot k \log_2(n/k)$ (see Section 2.1). Thus the contour set for $k/n = 0.25$ appears as it was "shifted" to the right with respect to the case $k/n = 0.15$ in Figure 3. As for the quantization of the projections, provided that the number of measurements is compatible with the sparsity level as

explained before, we can observe that the value of NMSE_R decreases as $\mathrm{SNR}_{\mathbf{y}}$ becomes larger. In a practical scenario, the quantization step size Δ should be chosen in such a way to attain $\mathrm{SNR}_{\mathbf{y}} \geq 25$ dB, in order to be robust with the choice of m, which depends on the actual sparsity of the tampering and on the constant C and is therefore unknown at the CP side. In our experiments in the rest of the paper, we have set $C = 1.3$.

6. Rate Allocation

In Section 3 we have shown that the correlation model between the original and the tampered random projections can be written as

$$\widetilde{\mathbf{y}} = \mathbf{y} + \mathbf{b}. \tag{18}$$

Hereafter, we assume that $\mathbf{y}$ and $\mathbf{b}$ are statistically independent. This is reasonable if the tampering is considered independent from the original audio content.

Let $j = 1, \ldots, J$ denote the bitplane index and R_j the bitrate (in bits/symbol) needed to decode the jth bitplane. As mentioned in Section 3, the probability density function of $\mathbf{y}$ and $\mathbf{b}$ can be well approximated to be zero mean Gaussian, respectively, with variance σ_y^2 and σ_b^2. The rate estimation algorithm receives in input the source variance σ_y^2, the correlation noise variance σ_b^2, the quantization step size Δ, and the number of bitplanes to be encoded J and returns the average number of bits needed to decode each bitplane R^j, $j = 1, \ldots, J$. The value of σ_y^2 can be immediately estimated from the random projections at the time of hash generation. The value of σ_b^2 is set to be equal to the maximum MSE distortion between the original and the tampered signal, for which tampering identification can be attempted.

The rate allocated to each bitplane is given by

$$R^j = H\left(\mathbf{y}^j \mid \widetilde{\mathbf{y}}, \mathbf{y}^{j-1}, \mathbf{y}^{j-2}, \ldots, \mathbf{y}^1\right) \left[\frac{\text{bits}}{\text{sample}}\right] + \Delta R, \tag{19}$$

where $\mathbf{y}^j$ denotes the jth bitplane of $\mathbf{y}$. In fact LDPC decoding of bitplane j exploits the knowledge of the real-valued side information $\widetilde{\mathbf{y}}$ as well as previously decoded bitplanes $\mathbf{y}^{j-1}, \mathbf{y}^{j-2}, \ldots, \mathbf{y}^1$. Since we use nonideal channel codes with a finite sequence length m to perform source coding a rate overhead of approximately $\Delta R = 0.1$ [bit/sample] is added. The integral needed to compute the value of the conditional entropy in (19) is factored out in detail in our previous work [33].

7. 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.

Table 1: Perceptual SNR, sparsity factor k/n in the most "sparsifying" basis (in parentheses) and m/n ratio for the three considered tampering example.

	SNR_P [dB]	Sparsity (k/n)	m/n
T	20.3	9% (1D-DCT)	0.54
F	11.5	26% (2D-DCT)	0.66
TF	14.5	6% (Haar)	0.54

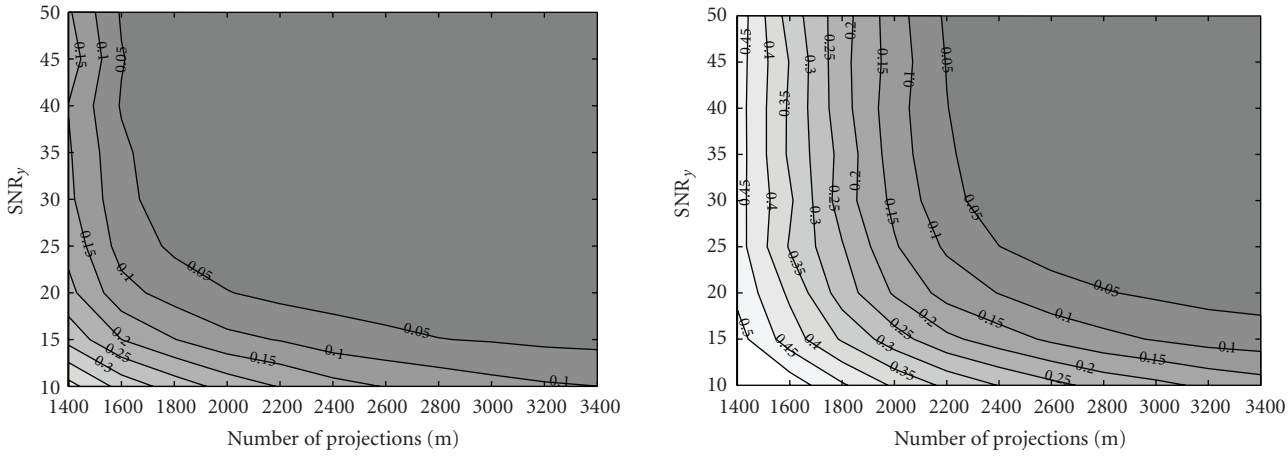
Time Localized Tampering (T). We have replaced some words in the speech at different positions, for a total tampering length of 3.75 seconds (about 11.7% of the total length of the audio sequence).

Frequency Localized Tampering (F). A low-pass phone-band filter (cut-off frequency at 3400 Hz and stop frequency at 4000 Hz) is applied to the entire original audio stream.

Time-Frequency Localized Tampering (TF). A cough at the beginning of the stream and the noise of the key ring in the middle are canceled out using the standard noise removal tool of the "Audacity" free audio editing software [34]. The noise removal tool implemented in this application is an adaptive filter, whose frequency response depends on the local frequency characteristics of the noise. In this case, the total time length of the attack is 4.36 seconds.

The reconstruction of the tampering has been attempted in 3 different bases, besides the log-energy domain: 1D DCT (discrete cosine transform across frequency bands of the same frame; this corresponds to extracting Mel frequency cepstral coefficients), 2D DCT (across time and frequency), and 2D Haar wavelet. Table 1 summarizes the perceptual SNRs and the sparsity of the three tampering examples, in the domain where its values is the lowest. It also reports the number of computed projections m in terms of the ratio m/n. Note that this ratio is always less than one (i.e., $m < n$), thus the adopted setting is coherent with the compressive sensing framework explained in Section 2.1. In the following, we evaluate two aspects of the system, namely: (1) the rate spent for Wyner-Ziv encoding the hash with respect to the rate that would have been spent for encoding and transmitting the projections without DSC; (2) the relation between the ℓ_1 and the inverse tangent norms of the quality of the reconstructed tampering in different domains.

7.1. 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



(a) $k/n = 0.15$ (b) $k/n = 0.25$

Figure 3: Normalized MSE of the reconstructed tampering as a function of the number of measurements m and the measures signal-to-quantization noise ratio SNR_y, expressed in dB.

hash direct quantization and transmission, that is, without using DSC (NO-WZ). Figure 4 depicts 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

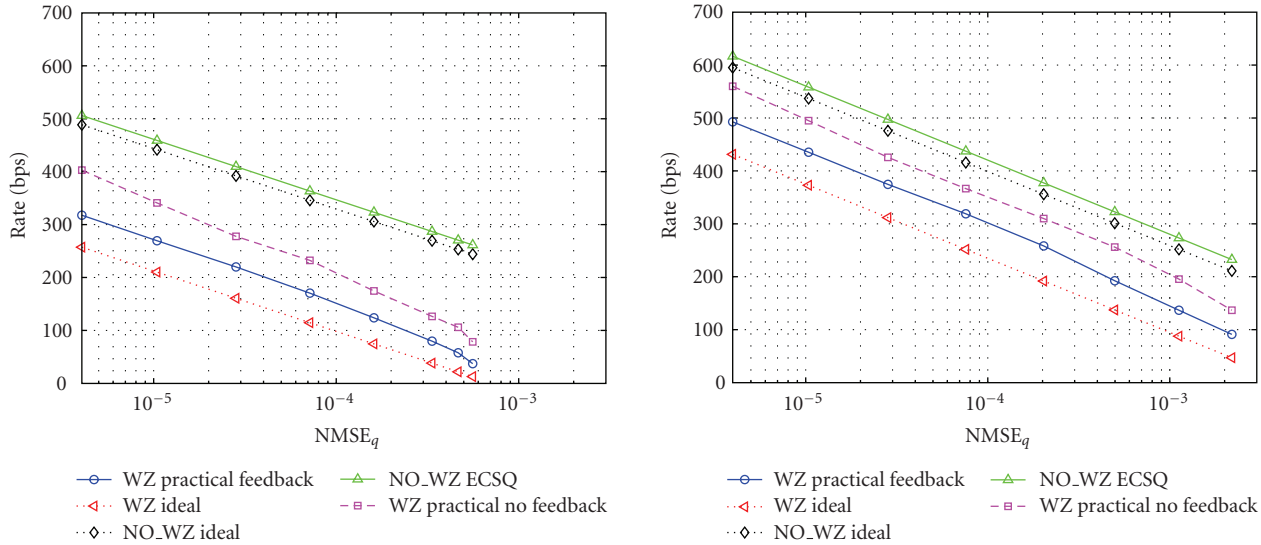
$$\text{NMSE}_q = \frac{||\mathbf{y} - \hat{\mathbf{y}}||_2^2}{\|\mathbf{y}\|_2^2}. \tag{20}$$

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).

We can make two main comments on the curves in the two graphs of Figure 4. The first difference between the frequency and the time tampering is that all the rate-distortion functions in the frequency attack are shifted upwards to higher rates, and have a steeper descending slope as the distortion increases. This is due to the fact that the frequency manipulation has a higher sparsity coefficient k/n, that is, more measurements are needed for signal reconstruction. Although in the real application no guess about the sparsity of the tampering can be made at the CP side, here we have fixed a different sparsity for the two kinds of attacks, in order to visually prove the effect of the number of measures on the hash length. Thus, even if the rate per measurement is the same in both the cases (it only depends on the signal energy, as expressed in (5) and (7)), the rate in bits per second has slopes and offsets proportional to the number of measurements m. Clearly, if we did not use compressive sensing to reduce the dimensionality of the data (i.e., $\mathbf{y} = \mathbf{x}$ in our setting), the rate required for the hash would have been equivalent to using random projections with $m = n$; therefore, the rate saving due to compressive sensing is approximately equal to the ratio m/n. The second interesting remark that emerges from Figure 4 is the different gap between the family of WZ rates (ideal, with feedback and without feedback) and the NO-WZ curves. As (8) suggests, the coding gain from NO-WZ to WZ strongly depends on the energy of the tampering, that is, to SNR_P (see Table 1). In the case of time attack, we have $\text{SNR}_P^T = 20.3$ dB, while $\text{SNR}_P^F = 11.5$ dB, thus according to (8) the bit saving achieved with WZ is smaller in the case of the frequency attack. As can be inferred from the graphs, this gain ranges from 20% to 70%.

7.2. 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. As described in Section 4.2, we are ultimately interested in finding the sparsest tampering representation. This requires in practice to evaluate the sparsity of the tampering in each basis expansion; we use for this purpose the inverse-tangent norm defined in (15). To validate the choice of this norm, we compare the optimal basis expansion predicted from the ℓ_1 norm and the inverse tangent norm with the actual best basis in terms of ℓ_2 reconstruction quality.

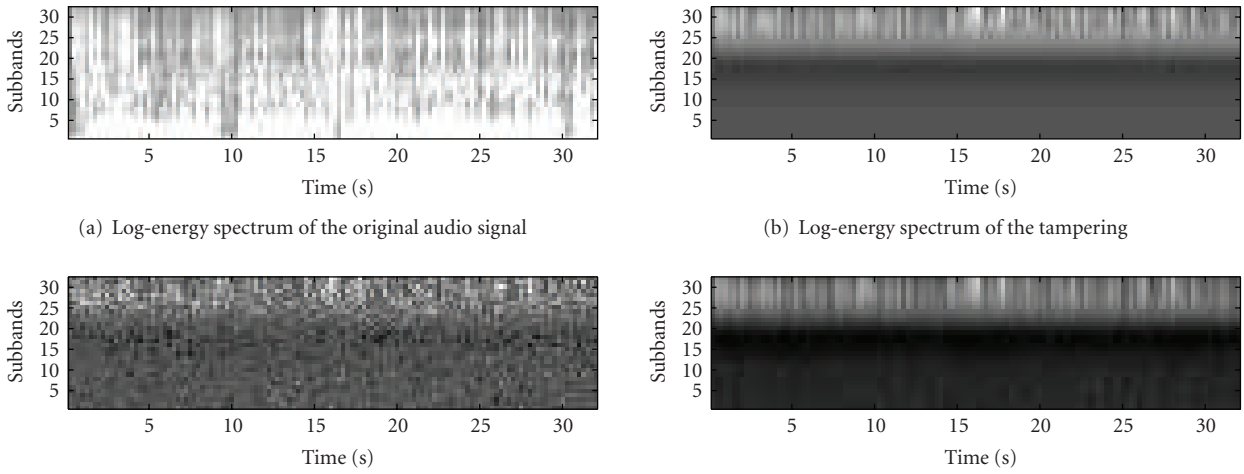
We evaluate the goodness of the tampering estimation by calculating the reconstruction normalized MSE between the



(a) Time sparse tampering, with a sparsity factor k/n set to 0.15

(b) Frequency sparse tampering, with sparsity factor $k/n = 0.25$

FIGURE 4: Rate-distortion function of the hash signature with different encoding approaches.



(a) Log-energy spectrum of the original audio signal

(b) Log-energy spectrum of the tampering

(c) Reconstructed tampering in log-energy domain

(d) Reconstructed tampering in 2D-DCT domain

FIGURE 5: Example of frequency tampering. The hash length is 200 bps.

log-energy spectrum of the original tampering and the log-energy spectrum of the estimated one

$$\mathrm{NMSE}_R = \frac{\|\hat{\mathbf{e}} - \mathbf{e}\|_2^2}{\|\mathbf{e}\|_2^2}. \tag{21}$$

Reconstruction NMSE values obtained with a fixed bit-rate for the hash are shown in Tables 2 (for 200 bps) and 3 (for 400 bps). The bit rate depends on the number of measurements m (given in Table 1) and on the number of bitplanes per measurement J. For a resulting rate of 200 bps, the number of bitplanes for the three kinds of attack (T, F, TF) is, respectively, 7, 5, and 6. When the rate is 400 bps, we have $J = 10$ for the time attack, $J = 8$ for the frequency attack, and $J = 9$ for the time-frequency tampering. From the tables it is clear that, by looking for a sparse tampering in other bases besides the canonical one (log-energy), better results can be achieved using the same hash length, as highlighted by the bold numbers in the tables. In particular, it can be observed that the wide-band, time-localized tampering is better reconstructed using the 1D-DCT basis, which is able to capture tampering correlations only along the frequency axis, avoiding tampering discontinuities over time. The frequency-localized tampering is better reconstructed using the 2D-DCT basis, due to its time extension and wide-band characterization which exhibits only a single discontinuity along the frequency axis. Finally, Haar wavelet is a good compromise to detect time-frequency localized tampering because it is able to deal with discontinuities along both time and frequency axes.

Table 2: NMSE_R for tampering reconstruction with a hash at a bit rate of 200 bps.

	Log-energy	1D-DCT	2D-DCT	Haar wavelet
T	$7.1\cdot10^{-3}$	$\mathbf{4.8\cdot10^{-3}}$	$2.5\cdot10^{-2}$	$7.9\cdot10^{-3}$
F	$1.1\cdot10^{-1}$	$3.6\cdot10^{-2}$	$\mathbf{8.6\cdot10^{-3}}$	$1.3\cdot10^{-2}$
TF	$2.3\cdot10^{-3}$	$1.5\cdot10^{-3}$	$4.3\cdot10^{-3}$	$\mathbf{1.4\cdot10^{-3}}$

Table 3: NMSE_R for tampering reconstruction with a hash at a bit rate of 400 bps.

	Log-energy	1D-DCT	2D-DCT	Haar wavelet
T	$2.4\cdot10^{-4}$	$\mathbf{2.1\cdot10^{-4}}$	$1.6\cdot10^{-2}$	$4.5\cdot10^{-4}$
F	$9.3\cdot10^{-2}$	$1.2\cdot10^{-2}$	$\mathbf{1.9\cdot10^{-3}}$	$3.1\cdot10^{-3}$
TF	$4.7\cdot10^{-5}$	$6.0\cdot10^{-5}$	$1.1\cdot10^{-3}$	$\mathbf{4.5\cdot10^{-5}}$

Table 4: ℓ_1-norm of the tampering using a fixed bit-rate for the hash signature of 200 bps.

	Log-energy	1D-DCT	2D-DCT	Haar wavelet
T	265.33	**183.06**	366.05	248.26
F	1219.12	1005.34	**251.41**	488.08
TF	509.88	**256.00**	445.95	260.71

Table 5: ℓ_1-norm of the tampering using a fixed bit-rate for the hash signature of 400 bps.

	Log-energy	1D-DCT	2D-DCT	Haar wavelet
T	344.57	**246.97**	543.06	338.81
F	1761.76	1394.95	**445.42**	731.58
TF	594.46	330.28	639.89	**325.64**

Tables 4 and 5 show the ℓ_1 norms of the reconstructed tampering coefficients in the four analyzed bases. Note that at a rate equal to 200 bps, the ℓ_1 norm suggests, for the time-frequency (TF) tampering, that the best reconstruction is with the 1D-DCT coefficients. However, Table 2 indicates that the best reconstruction is actually in the Haar wavelet domain. This is due to the noise introduced by compressive sensing recovery at low rates, which makes the use of the ℓ_1 norm as an estimator of the sparsity more error-prone. This effect is partially alleviated using the inverse tangent norm, as shown in Tables 6 and 7.

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.

8. Conclusions

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. In practice, such an approach is feasible only if the bit length of the hash is not too large; we have found that encoding the hash signature through a distributed source coding paradigm enables a consistent reduction of the transmitted bits, especially when the strength of the tampering is small compared to the original signal energy. The hash size may be further decreased in the future by considering weighted ℓ_1 minimization [32] to reduce the number of measurements required by the algorithm.

Table 6: Inverse tangent norm of the tampering using a fixed bit-rate for the hash signature of 200 bps.

	Log-energy	1D-DCT	2D-DCT	Haar wavelet
T	270.36	**166.68**	455.10	252.26
F	1115.59	793.15	**187.00**	323.35
TF	324.44	150.30	349.18	**136.60**

Table 7: Inverse tangent norm of the tampering using a fixed bit-rate for the hash signature of 400 bps.

	Log-energy	1D-DCT	2D-DCT	Haar wavelet
T	324.41	**224.92**	675.91	334.40
F	1586.11	1087.59	**412.81**	575.68
TF	308.54	196.59	536.68	**171.91**

Acknowledgment

This work has been partially sponsored by the EU under Visnet II Network of Excellence.

References

[1] M. Steinebach and J. Dittmann, "Watermarking-based digital audio data authentication," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 10, pp. 1001–1015, 2003.

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

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

[4] D. Kundur and D. Hatzinakos, "Digital watermarking for telltale tamper proofing and authentication," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1167–1180, 1999.

[5] R. Tu and J. Zhao, "A novel semi-fragile audio watermarking scheme," in *Proceedings of the 2nd IEEE Internatioal Workshop on Haptic, Audio and Visual Environments and Their Applications (HAVE '03)*, pp. 89–94, Ottawa, Canada, September 2003.

[6] P. W. Wong, "A public key watermark for image verification and authentication," in *Proceedings of IEEE International Conference on Image Processing (ICIP '98)*, vol. 1, pp. 455–459, Chicago, Ill, USA, October 1998.

[7] M. U. Celik, G. Sharma, E. Saber, and A. M. Tekalp, "Hierarchical watermarking for secure image authentication with localization," *IEEE Transactions on Image Processing*, vol. 11, no. 6, pp. 585–595, 2002.

[8] C.-S. Lu, S.-K. Huang, C.-J. Sze, and H.-Y. M. Liao, "Cocktail watermarking for digital image protection," *IEEE Transactions on Multimedia*, vol. 2, no. 4, pp. 209–224, 2000.

[9] C.-S. Lu, H.-Y. M. Liao, and L.-H. Chen, "Multipurpose audio watermarking," in *Proceedings of the 15th International Conference on Pattern Recognition (ICPR '00)*, vol. 3, pp. 282–285, Barcelona, Spain, September 2000.

[10] M. K. Mıhçak and R. Venkatesan, "A perceptual audio hashing algorithm: a tool for robust audio identification and information hiding," in *Proceedings of the 4th International Workshop on Information Hiding (IH '01)*, vol. 2137, pp. 51–65, Pittsburgh, Pa, USA, April 2001.

[11] J. Haitsma and T. Kalker, "A highly robust audio fingerprinting system with an efficient search strategy," *Journal of New Music Research*, vol. 32, no. 2, pp. 211–221, 2003.

[12] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A review of algorithms for audio fingerprinting," in *Proceedings of IEEE Workshop on Multimedia Signal Processing (MMSP '02)*, pp. 169–173, St. Thomas, Virgin Islands, USA, December 2002.

[13] S. Roy and Q. Sun, "Robust hash for detecting and localizing image tampering," in *Proceedings of the 14th IEEE International Conference on Image Processing (ICIP '07)*, vol. 6, pp. 117–120, San Antonio, Tex, USA, October 2007.

[14] Y.-C. Lin, D. Varodayan, and B. Girod, "Image authentication based on distributed source coding," in *Proceedings of IEEE International Conference on Image Processing (ICIP '07)*, vol. 3, pp. 5–8, San Antonio, Tex, USA, September-October 2007.

[15] Y.-C. Lin, D. Varodayan, and B. Girod, "Spatial models for localization of image tampering using distributed source codes," in *Proceedings of the International Picture Coding Symposium (PCS '07)*, Lisbon, Portugal, November 2007.

[16] K. Chono, Y.-C. Lin, D. Varodayan, Y. Miyamoto, and B. Girod, "Reduced-reference image quality estimation using distributed source coding," in *Proceedings of IEEE International Conference on Multimedia and Expo (ICME '08)*, Hannover, Germany, June 2008.

[17] D. Varodayan, Y.-C. Lin, and B. Girod, "Audio authentication based on distributed source coding," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '08)*, pp. 225–228, Las Vegas, Nev, USA, March-April 2008.

[18] M. Tagliasacchi, G. Valenzise, and S. Tubaro, "Localization of sparse image tampering via random projections," in *Proceedings of the 15th IEEE International Conference on Image Processing (ICIP '08)*, pp. 2092–2095, San Diego, Calif, USA, October 2008.

[19] E. J. Candès, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians (ICM '06)*, Madrid, Spain, August 2006.

[20] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Processing Magazine*, vol. 24, no. 4, pp. 118–121, 2007.

[21] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling: a sensing/sampling paradigm that goes against the common knowledge in data acquisition," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 21–30, 2008.

[22] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973.

[23] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976.

[24] B. Girod, A. M. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.

[25] V. K. Goyal, A. K. Fletcher, and S. Rangan, "Compressive sampling and lossy compression: do random measurements provide an efficient method of representing sparse signals?" *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 48–56, 2008.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[27] E. van den Berg and M. P. Friedlander, "In pursuit of a root," Tech. Rep. TR-2007-19, Department of Computer Science, University of British Columbia, Vancouver, Canada, June 2007, preprint , http://www.optimization-online.org/DB_FILE/2007/06/1708.pdf.

[28] D. Varodayan, A. Aaron, and B. Girod, "Rate-adaptive codes for distributed source coding," *Signal Processing*, vol. 86, no. 11, pp. 3123–3130, 2006.

[29] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: architecture, techniques and evaluation," in *Proceedings of the International Picture Coding Symposium (PCS '07)*, vol. 6, pp. 14496–14410, Lisbon, Portugal, November 2007.

[30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley & Sons, New York, NY, USA, 1991.

[31] L. Rabiner and B. H. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, Upper Saddle River, NJ, USA, 1993.

[32] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[33] R. Bernardini, M. Naccari, R. Rinaldo, M. Tagliasacchi, S. Tubaro, and P. Zontone, "Rate allocation for robust video streaming based on distributed video coding," *Signal Processing: Image Communication*, vol. 23, no. 5, pp. 391–403, 2008.

[34] Audacity, http://audacity.sourceforge.net.

International Journal of Digital Multimedia Broadcasting

Special Issue on
Multicell Cooperation and MIMO Technologies for Broadcasting and Broadband Communications

Call for Papers

The wireless industry is experiencing an unprecedented increase in the number and sophistication of broadcasting and broadband communication systems. The growing diffusion of new services, like mobile television and multimedia communications, emphasizes the need of advanced transmission techniques that can fundamentally increase the system capacity. In this context, the multicell collaborative transmission is becoming one of a major subject of research in the wireless communication community as it has been identified as one of the underlying principles for future wireless communication systems. Further, if perfect cooperation is assumed, it allows the entire network to be viewed as a single cell MIMO system with a distributed antenna array at the base station.

This special issue aims at promoting state-of-the-art research contributions from all research areas either directly involved in or contributing to improving the issues related to multicell cooperation and MIMO technologies for broadcasting and broadband communications. Topics of interest of this special issue include but are not limited to:

- Information theoretic aspects of cooperative communication systems
- Cooperative broadcasting with uninformed transmitter
- Effects of partial and incomplete channel state information in cooperative/MIMO systems
- Advances in MIMO and MISO algorithms and applications
- Physical and MAC layer issues in cooperative/MIMO communications
- Performance analysis of distributed MIMO techniques
- Space-time diversity technologies and space-time coding
- Practical implementations, test-beds, and demonstrations
- Standardization and deployment in 3G+, 4G, and beyond

Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://www.hindawi.com/journals/ijdmb/guidelines.html. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/, according to the following timetable:

Manuscript Due	September 1, 2009
First Round of Reviews	December 1, 2009
Publication Date	March 1, 2010

Lead Guest Editor

Hongxiang Li, Department of Electrical and Computer Engineering, North Dakota State University, Fargo, ND 58103, USA; hongxiang.li@ndsu.edu

Guest Editors

Jinyun Zhang, Mitsubishi Electrical Research Labs (MERL), Cambridge, MA 02139, USA; jzhang@merl.com

Lingjia Liu, Samsung Electronics, Richardson, TX 75082, USA; lliu@sta.samsung.com

Guoqing Li, Intel Corporation, Hillsboro, OR 97124, USA; guoqing.c.li@intel.com

Younsun Kim, Department of Electrical Engineering, University of Washington, Seattle, WA 98195, USA; younsun@u.washington.edu

EURASIP Journal on Wireless Communications and Networking

Special Issue on

Signal Processing-Assisted Protocols and Algorithms for Cooperating Objects and Wireless Sensor Networks

Call for Papers

With the advent of the so-called Internet of Things (IoTs), we will witness an unprecedented growth in the number of networked terminals and devices. In attaining this IoT vision, a class of energy- and, in general, resource-constrained systems like Wireless Sensor Networks (WSNs), networks of cooperating objects and embedded devices such as RFIDs, or networks for Device-to-Device (D2D) and Machine-to-Machine (M2M) communications are to play a fundamental role. The paradigm shift from general-purpose data networks to application-oriented networks (e.g., for parameter or random field estimation, event detection, localization, and tracking) clearly calls for further optimization at the physical, link, and network layers of the protocol stack. Interestingly, the above-mentioned estimation/detection/localization/tracking problems have been addressed for years by the signal processing community, this resulting into a number of well-known algorithms. Besides, some inspiration could be also borrowed from other communication schemes, such as MIMO and beamforming techniques or cooperative communications that were traditionally developed for wireless data networks, or even from other fields such as mathematical biology (e.g., networks of coupled oscillators). However, the challenge now is to enhance such algorithms and schemes and make them suitable for decentralized and resource-constrained operation in networks with a potentially high number of nodes. Complementarily, the vast literature produced by the information theory community, on the one hand, reveals the theoretical performance limits of decentralized processing (e.g., distributed source coding) and, on the other, offers insight on the scalability properties of such large networks and their behavior in the asymptotic regime. Realizing the information-theoretic performance with practical decentralized networking, radio resource management schemes, routing protocols, and other network management paradigms is a key challenge.

The objective of this Special Issue (whose preparation is carried out under the auspices of the EC Network of Excellence in Wireless Communications NEWCOM++) is to gather recent advances in the areas of cooperating objects, embedded devices, and wireless sensor networks. The focus is on how the design of future physical, link, and network layers could benefit from a signal processing-oriented approach. Specific topics for this Special Issue include but are not limited to:

- Decentralized parameter estimation
- Estimation of random fields
- Distributed MIMO and beamforming
- Decentralized and cooperative time and frequency synchronization
- Cooperative event detection
- Data gathering and data fusion
- Data-centric multihop techniques and routing
- Scalability and asymptotic laws for in-network distributed estimation/detection
- Energy-saving algorithms and protocols
- Feedback-limited scheduling and MAC protocols
- Decentralized joint source-channel coding
- Cooperative localization and tracking
- Topology control in resource-constrained networks
- Low-complexity opportunistic networking protocols

Before submission, authors should carefully read over the journal's Author Guidelines, which are located at http://www.hindawi.com/journals/wcn/guidelines.html. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/ according to the following timetable:

Manuscript Due	February 1, 2010
First Round of Reviews	May 1, 2010
Publication Date	August 1, 2010

Lead Guest Editor

Carles Antón-Haro, Centre Tecnològic de Telecomunicacions de Catalunya (CTTC), 08860 Castelldefels, Barcelona, Spain; carles.anton@cttc.es

Guest Editors

Davide Dardari, WiLAB, University of Bologna at Cesena, Cesena (FC), Italy; ddardari@ieee.org

Osvaldo Simeone, Center for Wireless Communications and Signal Processing Research, New Jersey Institute of Technology (NJIT), Newark, NJ, USA; osvaldo.simeone@njit.edu

Roberto Verdone, WiLAB at the University of Bologna, Bologna, Italy; roberto.verdone@unibo.it

Hindawi Publishing Corporation
http://www.hindawi.com

EURASIP Journal on Advances in Signal Processing

Special Issue on
Applications of Time-Frequency Signal Processing in Wireless Communications and Bioengineering

Call for Papers

Time-frequency signal processing is a well-established area with applications ranging from bioengineering and wireless communications to earthquake engineering and machine monitoring. Signals in these applications are typically nonstationary and as such require joint time-frequency analysis. The objective of this special issue is to bring together theoretical results and application of time-frequency methodologies from investigators in the wireless communications and bioengineering disciplines.

While novel theoretical results and applications of time-frequency signal processing in wireless communications and biomedical systems will be preferred, applications in other areas will also be considered. Likewise, this issue will emphasize methodologies related to Priestley's evolutionary spectrum and the fractional Fourier transform, but other methodologies will also be considered.

The intended focus of this issue will be on presenting time-frequency signal processing applications to wireless communications and biomedical systems using evolutionary spectral techniques and fractional Fourier transform.

Topics of interest include, but are not limited to:

- Biomedical systems: EEG, ECG waveforms and heart sound, vibroartrographic signals emitted by human knee joints, EEG signals, and various other biomedical waveforms analyzed by time-frequency techniques
- Wireless communications: time-frequency receivers, channel characterization, channel diversity, time-varying modulation schemes, and suppressing nonstationary interference as chirp jammers

Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://www.hindawi.com/journals/asp/guidelines.html. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/ according to the following timetable:

Manuscript Due	January 1, 2010
First Round of Reviews	April 1, 2010
Publication Date	July 1, 2010

Lead Guest Editor

Luis F. Chaparro, Department of Electrical and Computer Engineering, University of Pittsburgh, Pittsburgh, PA, USA; chaparro@ee.pitt.edu

Guest Editors

Aydin Akan, Department of Electrical and Electronics Engineering, Istanbul University, Istanbul, Turkey; akan@istanbul.edu.tr

Syed Ismail Shah, Department of Computing and Technology, Iqra University, Islamabad, Pakistan; ismail@iqraisb.edu.pk

Lutfiye Durak, Department of Electronics and Communications Engineering, Yildiz Technical University, Istanbul, Turkey; lutfiye@ieee.org