Audio Watermarking: A Survey

Mr. Jay R. Desai[1] **Mr. Mitul M. Patel**[2]
[2]Assistant Professor
[1,2]Department of Electronics and Communication Engineering
[1,2]Parul Institute of Engineering and Technology, Limda, Vadodara, India

Abstract — By virtue of the new advancements in computer and telecommunication networks, multimedia files are produced stored and distributed easily across the globe. However, the ownership and copyright of multimedia files are not usually protected. Digital watermarking has been proposed in recent years as a means of protecting multimedia contents from intellectual piracy .A perfect reproduction in digital domain has promoted the protection of intellectual ownership and the prevention of unauthorized tampering of multimedia data to become an important technological and research issue.

Digital watermarking has been proposed as a new, alternative method to enforce intellectual property rights and protect digital media from tampering. Digital watermarking is defined as imperceptible, robust and secure communication of data related to the host signal, which includes embedding into and extraction from the host signal. The main challenge in digital audio watermarking is that if the perceptual transparency parameter is fixed, the design of a watermark system cannot obtain high robustness and a high watermark data rate at the same time.

I. INTRODUCTION:

Multimedia information hiding (MIH) techniques have aimed to help to preserve and authenticate the values of multimedia information such as text, digital-audio, images, and video, help to hide imperceptible marks such as copyright notice into them, or even help to prevent their unauthorized copying. MIH techniques are, in general, composed of content protection of multimedia information such as watermarking and steganography that means hiding multimedia information in other multimedia information. Since it is possible to use MIH techniques together with cryptographic techniques, they are applicable for secure content authentication such as fingerprint.

Globalization and internet are the two main reasons for the growth of research and spreading of information and hence they become the greatest tool for malicious user to attack and pirate the digital media. The watermarking technique during the evolution was used on images, and is termed as Image Watermarking. In Image watermarking, text or any another image is embedded into original image for ownership protection. But as time goes the malicious user has started to extract the watermark and creating challenges for the developers [1]. Thus, developers have found another digital embedding media as audio and termed such watermarking as Audio Watermarking. In Audio watermarking, text, image or audio is embedded into original audio signal. It is very difficult to secure digital information particularly the audio and audio watermarking has become a challenge to developers because of the impression it has created in preventing copyrights and ownership of the music. Digital watermarking is a technique by which ownership information is embedded into the host signal in a way that the embedded information is imperceptible, and robust against intentional and unintentional attacks [2]. Audio watermarking has aimed to embed codes to protect the copyright in audio content that are inaudible to and inseparable by users, and to detect embedded codes from watermarked signals.

General audio digital watermark can be divided into time-domain and the transformation-domain methods [2]. Time domain method is that embedded watermark on amplitude of the samples about original audio signal, it is a simple method to be realized, and can embedded in a larger volume of data, but less robust. Because of the wavelet transform has a very good performance of compressed energy, in the transform domain of audio watermarking method, embedded watermark to the strong energy signal component, for this we can further enhance the robustness, so it is widely used.

In general, audio watermarking methods must satisfy three requirements to provide a useful and reliable form of copyright protection: (a) inaudibility (inaudible to humans with no sound distortion caused by the embedded data), (b) confidentiality (secure and undetectable concealment of embedded data), and (c) robustness (not affected when subjected to techniques such as data compression).The first requirement (inaudibility) is the most important in the method of audio watermarking because this must not affect the sound quality of the original audio. If the sound quality of the original is degraded, the original content may lose its commercial value. The second requirement (confidentiality) is important to conceal watermarks to protect copyright, and it is important that users do not know whether the audio content contains watermarking or not. The last requirement (robustness) is important to ensure the watermarking methods are tamper-proof to resist any manipulations by illegal users.

II. DIGITAL WATERMARKING APPLICATIONS:

There are various applications in which watermarking can be used. Some of the applications are mentioned below.

A. *Ownership Protection And Proof Of Ownership[4]*

In ownership protection application, the watermark embedded signal contains a proof of ownership. The embedded information is robust and secure against various attacks and can be extracted in a case of dispute of ownership. There can be the situations where some other person claims that it is his own [3]. For example, „x‟ has created an audio and put it on website with copyright notice. An attacker „y‟ then steals the audio file and modified the copyright notice and claim for ownership. So, this situation can be avoided by registering audio to the Copyright office [4]. But registering to copyright office is too costly. So another method is to embed copyright information into the audio file. So when such worst situation arises at that time owner of the audio file can take the dispute copy and original copy of audio file and then recovering the watermark, which contains information regarding owner, from the dispute copy and can proved the ownership[5].

B. Device Control [4]

Device control is a technique in which device can be control by identifying the watermark from the host signal. For example a unique identifier code is given into printed and distributed images such as magazine advertisements, packaging, tickets, and so on. After the image is captured by a mobile phone's camera, the watermark is read by the software on the phone and the identifier is used to direct a web browser to an associated web site. Similarly any audio, which contains the watermark, can be recorded by mobile and read by using software and then the identifier is used to direct a web browser to an associated web site [6].

C. Authentication and Tampering Detection [4]

In this application information is embedded in the host signal and can be used to check if the host signal is tampered. This situation is important because it is necessary to know about the tampering caused to the media signal. The tampering is sometime a cause of forging of the watermark which has to be avoided [8]. For example if the audio is divided into frames and watermarked is embedded into each frame then if some part of audio file is modified then it shows that the audio file is tampered. It is also useful for authentication [9]. For example, authentication can be useful in a police investigation of a crime. Let the police receive a surveillance audio that has been tampered with. If the audio is authenticated with a traditional signature, all the police would know that the audio is inauthentic and cannot be trusted. However, if they use a watermark for authentication, they might discover that each frame of the audio is reliable except for some part of an audio. This would be strong evidence that the identity of someone involved in the crime was removed from the audio [6].

D. Copyright Protection [4]

Copyright protection is the most important application of watermarking. The objective is to embed information identifies the copyright owner of the digital media, in order to prevent other parties from claiming the copyright.

This application requires a high level of robustness to ensure that embedded watermark cannot be removed without causing a significant distortion in digital media. Additional requirements beside the robustness have to be considered. For example, the watermark must be unambiguous and still resolve rightful ownership if other parties embed additional watermarks.

E. Fingerprinting [4]

The objective of this application is to convey information about the legal recipient rather than the source of digital media, in order to identify single distributed copies of digital work. It is very similar to the serial number of software product. In this application a different watermark embedded into each distributed copy. In contrast the first application where only a single watermark is embedded into all copies of digital media. As well as copyright protection application of watermarking, fingerprinting requires high robustness.

F. Broadcast Monitoring [4]

Producers of advertisements or audio and video works want to make sure that their works are broadcasted on the time they purchase from broadcasters. The low-tech method of broadcast monitoring is to have human observers watch the broadcasting channels and record what they see or hear. This method is costly and error prone. The solution is to replace the human monitoring with automated monitoring. One method of automated broadcast monitoring is to use the watermarking techniques. With watermarking we can embed an identification code in the work being broadcasted. A computer-based monitoring system can detect the embedded watermark, to ensure that they receive all of the airtime they purchase from the broadcasters.

III. PROPERTIES OF DIGITAL WATERMARKING:

The watermarking system can be characterized by a number of properties. The important properties are as follows:

A. Perceptual Transparency

The goal of the any watermark embedding algorithm is to insert watermark data without changing the quality of the host signal and also the embedded watermark should not be audible or it should be inaudible to the human. So, the perceptual transparency is referred as similarity between original and embedded audio signal. It is also referred as Imperceptibility.

The quality of the watermarked audio may be changed, either intentionally or unintentionally before a person perceives it. In such a case, imperceptibility can be considered as a perceptual similarity between the watermarked audio and the original audio at the point at which they are presented to a consumer. So, the data embedded into the original audio signal through the watermarking should not produce annoying or any undesirable noise or in another words watermarked audio should be indistinguishable from the original audio otherwise it loses its commercial value[4].

B. Embedding Effectiveness [4]

The effectiveness of a watermarking system is the probability that the output of the embedded will be watermarked. The cover work is said to be watermarked when input to a detector result in positive detection. The effectiveness of a watermarking system may be determined analytically or empirically by embedding a watermark in a large number of cover works and detect the watermark. The percentage of cover works that result in positive detection will be the probability of effectiveness.

C. Fidelity [4]

In general, the fidelity of a watermark system refers to the perceptual similarity between the original and the watermarked version of the cover work. However, watermarked work may be degraded in the transmission process prior to its being perceived by a person, a different definition of fidelity may be more appropriate. We may define watermarking system fidelity as a perceptual similarity between the un-watermarked and watermarked works at the point at which they are presented to a viewer.

D. *Data Payload [4]*

Data payload refers to the number of bits a watermark embeds in a unit of time or works. For audio, data payload refers to the number of embedded bits per second that are transmitted. Different applications require different data payload. For example, Copy control applications may require a few bits embedded in cover works.

E. *Blind or Informed Detector [4]*

We refer to the detector that requires the original, un-watermarked work as an informed detector. Informed detectors may require information derived from the original work rather than original work itself. Conversely, detectors that do not require the original work are referred to as blind detectors. Informed detector has a good performance in watermark extraction. However, this will result in a huge number of original works have to be stored.

F. *False Positive Rate [4]*

A false positive is the detection of a watermark in a cover work that does not actually contain one. When we talk of a false positive rate, we refer to the number of false positives we expect to occur in a given number of runs of the detector.

G. *Robustness, Security and Cost [4]*

Robustness refers to the ability to detect the watermark after common signal processing operations. Audio watermarking needs to be robust to temporal filtering, A/D conversion, time scaling, etc. not all applications of watermarking require all the forms of robustness. This depends on the nature of application of watermarking system. The security of a watermark refers to its ability to resist hostile attacks. Hostile attack is the process specifically intended to thwart the watermark's purpose. The types of attacks can fall in three categories: unauthorized removal, unauthorized embedding, and unauthorized detection. The Cost of watermarking system refers to the speed with which embedding and detection must be performed and the number of embedders and detectors that must be deployed. Other issues include the whether the detector and embedder are to be implemented as hardware device or as software application or plug-ins.

IV. DIFFERENT TRANSFORMATION TECHNIQUES OF AUDIO WATERMARKING:

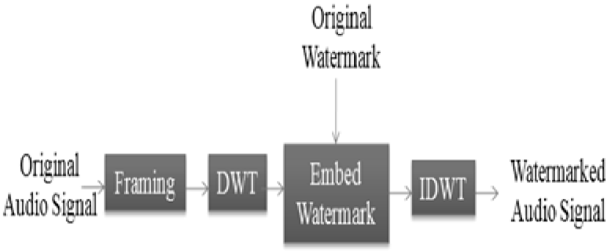


Fig. 1: Discrete Wavelet Transform.

There are mainly two transformation techniques available for audio watermarking:

A. *DISCRETE WAVELET TRANSFORM:*

1) Embedding Algorithm:
2) Extracting Algorithm

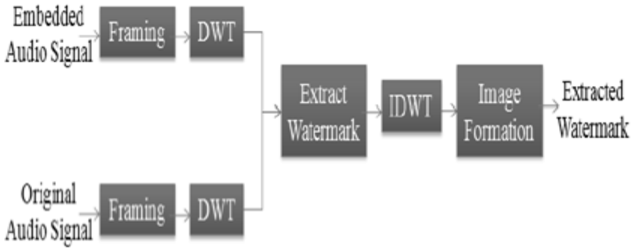


Fig. 2: Discrete Cosine Transform.

1) *Theory:*

Discrete Wavelet Transform (DWT) is any wavelet transform for which the wavelets are discretely sampled. Key advantage it has over Fourier transforms is temporal resolution: it captures both frequency and location information.

2) *One level DWT*

The operation of 1-level DWT decomposition is to separate high pass and low pass components. So, process involves passing the time-domain signal x[n] through a high pass filter h[n] and down sampling the signal obtained yields detailed coefficients (D) and passing x[n] through low pass filters g[n] and also down sampling generated approximate coefficients (A).

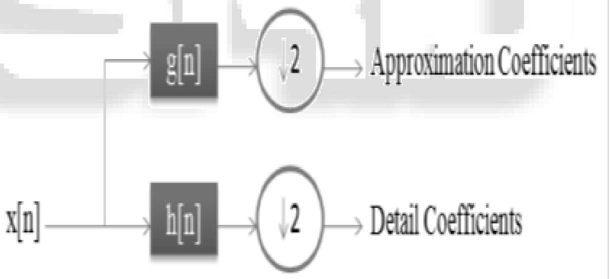


Fig. 3: Block Diagram of 1-Level DWT

By mathematically, the signal x[n] is passed through the low pass filter whose impulse response is g[n].So, result can be generated by convolution of x[n] and g[n].

Ylow[n]=(x*g)[n]=∑_(x=-∞)^∞▒〖x[k]g(n-k)〗

Same way the signal x[n] is passed through the high pass filter whose impulse response is h[n] and its result can be generated by convolution of x[n] and h[n].

Yhigh[n]=(x*h)[n]=∑_(x=-∞)^∞▒〖x[k]h(n-k)〗

Two filters are related to each other and they are known as a quadrature mirror filter.

Due to half frequencies removed from the signal, half the samples discarded according to Nyquist‟s rule. So, the filter outputs are down sampled by 2. At last, the approximate and detail coefficients of the signal can be calculated by following way:

Ylow[n]= ∑_(x=-∞)^∞▒〖x[k]g(2n-k)〗

Yhigh[n]= ∑_(x=-∞)^∞▒〖x[k]h(2n-k)〗

This decomposition has the time resolution since only half of each filter output characterises the complete signal. However, each output has half the frequency band of the input so that the frequency resolution has also been doubled.

B. Multi-level DWT

If we want to further increase frequency resolution the decomposition procedure is repeted. The approximation coefficients are decomposed with low and high pass filters and then down sampled again.

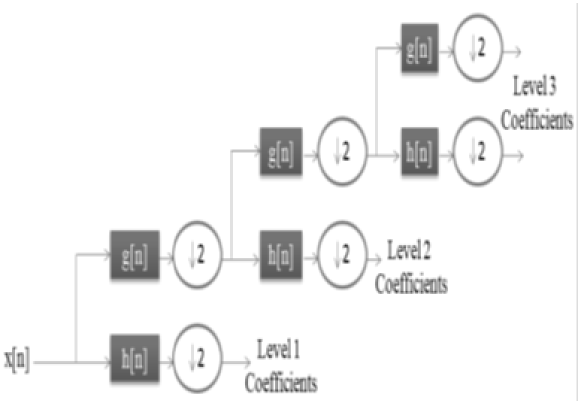


Fig. 4: Block Diagram Of 3-Level DWT [20]

Due to the signal decomposition process the input signal must be a multiple of 2 * n where n is the number of levels. At each signal decomposition level, the half band filters produce signals which having only half the frequency band. So, the frequency resolution becomes doubles. According to Nyquist rule, if the original signal has a highest frequency of ω, which requires a sampling frequency of 2*ω radians, then it now has a highest frequency of ω/2 radians. It can now be sampled at a frequency of ω radians thus if we discarding half the samples then also no loss of information and hence time resolution becomes half as the entire signal is now represented by only half the number of samples [4].

C. Inverse Discrete Wavelet Transform

The reconstruction of the original signal from the wavelet coefficients is the exact reverse procedure of the decomposition. Figure shows the reconstruction of the original signal from the wavelet coefficients. The approximation and detail coefficients at every level are up sampled by two, passed through the low pass and high pass synthesis filters and then added [7]. This process is continued through the same number of levels as in the decomposition process to obtain the original signal.

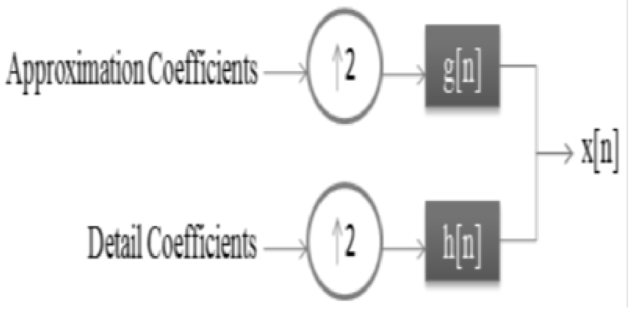
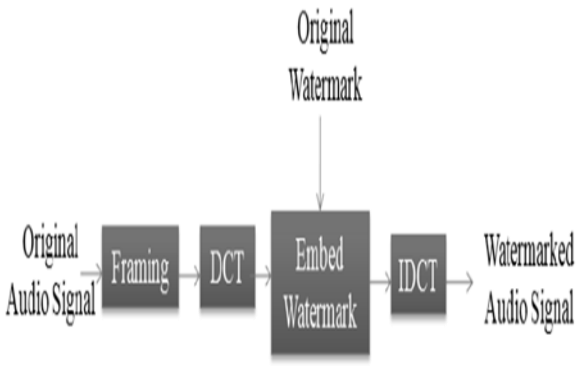


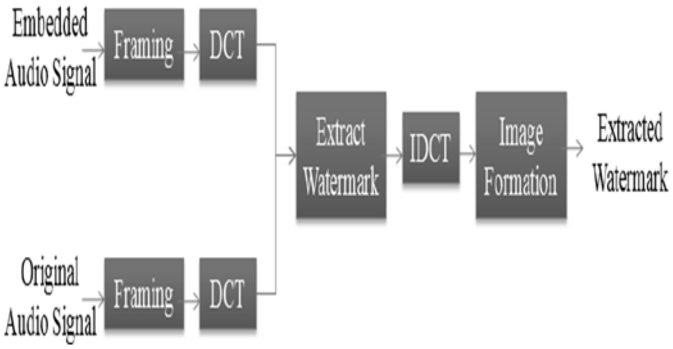
Fig. 5: Block Diagram Of 1-Level IDWT [7]

D. DISCRETE COSINE TRANSFORM:

1) Embedding Algorithm:



2) Extracting Algorithm:



E. Theory:

DCT expresses a finite sequence of data points in terms of a sum of cosine functions oscillating at different frequencies. They are important to numerous applications in science and engineering, from lossy compression of audio and images, to spectral methods for the numerical solution of partial differential equations.

The use of cosine functions is critical in these applications: for compression, it turns out that cosine functions are much more efficient, whereas for differential equations the cosines express a particular choice of boundary conditions.

There are several variants of the DCT with slightly modified definitions. The N real numbers x0, ..., xN-1 are transformed into the N real numbers X0, ..., XN-1 according to one of the formulas:

1) DCT-I

Xk = 1/2 (x_0+(-1)^k x_(N-1))+∑_(n=1)^(N-2)▒〖x_n cos[π/(N-1) nk] k=0,….,N-1〗

Some authors further multiply the x0 and xN-1 terms by √2, and correspondingly multiply the X0 and XN-1 terms by 1/√2. This makes the DCT-I matrix orthogonal, if one further multiplies by an overall scale factor of √2/√(N-1), but breaks the direct correspondence with a real-even DFT.

The DCT-I is exactly equivalent to a DFT of 2N-2 real numbers with even symmetry.

Thus, the DCT-I corresponds to the boundary conditions: xn is even around n=0 and even around n=N-1; similarly for Xk.

2) DCT-II

Xk = ∑_(n=0)^(N-1)▒〖x_n cos[π/N (n+1/2)k] k=0,….,N-1〗

The DCT-II is probably the most commonly used form, and is often simply referred to as "the DCT".

This transform is exactly equivalent (up to an overall scale factor of 2) to a DFT of 4N real inputs of even symmetry where the even-indexed elements are zero. That is, it is half of the DFT of the 4N inputs , where y_2n=0, for 0<=n<N, and for 0<n<2N Some authors further multiply the X0 term by 1/√2 and multiply the resulting matrix by an overall scale factor of (√2)⁄N.

The DCT-II implies the boundary conditions: xn is even around n=-1/2 and even around n=N-1/2; Xk is even around k=0 and odd around k=N.

3) DCT-III

Xk =1/2 x_0+ ∑_(n=0)^(N-1)▒ 〖x_n cos[π/N (n)(k+1/2)] k=0,….,N-1〗

Because it is the inverse of DCT-II (up to a scale factor, see below), this form is sometimes simply referred to as "the inverse DCT" ("IDCT").

Some authors further multiply the x0 term by √2 and multiply the resulting matrix by an overall scale factor of

The DCT-III implies the boundary conditions: xn is even around n=0 and odd around n=N; Xk is even around k=-1/2 and even around k=N-1/2.

4) DCT-IV

Xk = ∑_(n=0)^(N-1)▒ 〖x_n cos[π/N (n+1/2)(k+1/2)] k=0,….,N-1〗

The DCT-IV matrix becomes orthogonal (and thus, being clearly symmetric, its own inverse) if one further multiplies by an overall scale factor of (√2)⁄N

A variant of the DCT-IV, where data from different transforms are overlapped, is called the modified discrete cosine transform (MDCT)

The DCT-IV implies the boundary conditions: xn is even around n=-1/2 and odd around n=N-1/2; similarly for Xk.

5) Inverse Transforms

Using the normalization conventions above, the inverse of DCT-I is DCT-I multiplied by 2/(N-1). The inverse of DCT-IV is DCT-IV multiplied by 2/N. The inverse of DCT-II is DCT-III multiplied by 2/N and vice versa.

6) Multidimensional DCT

Multidimensional variants of the various DCT types follow straightforwardly from the one-dimensional definitions: they are simply a separable product (equivalently, a composition) of DCTs along each dimension. The 2d DCT-II is given by the formula

Xk1,k2 = ∑_(n_1=0)^(N_1-1)▒ 〖(∑_(n_2=0)^(N_2-1)▒ 〖 x_n1n2 cos⁡[π/N_2 (n_2+1/2) k_2])cos⁡[π/N_1 (n_1+1/2) k_1]〗 〗

= ∑_(n_1=0)^(N_1-1)▒ 〖(∑_(n_2=0)^(N_2-1)▒ 〖 x_n1n2 cos⁡[π/N_1 (n_1+1/2) k_1])cos⁡[π/N_2 (n_2+1/2) k_2]〗 〗

V. CONCLUSION

All watermarking systems are designed to achieve one goal that is embedding a hidden robust watermark into digital media. These systems have to satisfy two conflicting requirements. First, watermark must be immune against intentional and unintentional removal. Second, watermarked signal should maintain a good fidelity, i.e. watermark must be perceptually undetectable. To accomplish this task, variety of techniques has been exploited, and different domains are involved to enhance a certain application of watermarking and/or improve fidelity and robustness of watermarked signal.

REFERENCES

[1]Mazdak Zamani, Azizah Bt Abdul Manaf, Rabiah Bt Ahmad, Akram M. Zeki, Pritheega Magalingam, “A Novel Approach for Audio Watermarking”, IEEE Journel 2009

[2]Manie Kansal, Gursharanjeet Singh, B V Kranthi, “DWT, DCT and SVD based Digital Image Watermarking”, 2012 International Conference on Computing Sciences @ 2012 IEEE.

[3]Mrs. Rashmi G.Dukhi, “Watermarking: A Copyright Protection Tool”, 3rd International Conference on Electronic technology, pp. 36-41, April-2011.

[4]A thesis submitted by Jian Wang, NEW DIGITAL AUDIO WATERMARKING ALGORITHMS FOR COPYRIGHT PROTECTION, M.Sc. Department of Computer Science b National University of Ireland, Maynooth, Co. Kildare, Ireland, Septmber 2011.

[5]Lawrence R. Rabiner and Ronald W. Schafer, Introduction to Digital Speech Processing, Foundation and trends in signal processing 1:1-2(2007)

[6]http://www.umiacs.umd.edu/~desin/Speech1/node10.html

[7]En.wikipedia.org/wiki/discrete_wavelet_transform

[8]Ho Shuet Mun, Ng Ling Mei, “Audio Watermarking Using Time-Frequency Compression Expansion”, Foo Say Wei IEEE 2004

[9]A Thesis Submitted to the Graduate Faculty of the Louisiana State by, Rajkiran Ravula Bachelor of Engineering in Electrical and Electronics Engineering, AUDIO WATERMARKING USING TRANSFORMATION TECHNIQUES, Osmania University, 2006 Hyderabad, India December, 2010

[10]Xiumei Wen, Xuejun Ding, Jianhua Li, Liting Gao, Haoyue Sun, “An Audio Watermarking Algorithm Based on Fast Fourier Transform”, 978-0-7695-3876-1/09 $25.00 © 2009 IEEE

[11]Dai Tracy Yang, Rade Petrovic, “Audio Watermarking In Compressed Domain”,Telsiks 2009

[12]Mangal Patil, J. S. Chitode, “Improved Technique For Audio Watermarking Based On Discrete Wavelet Transform”, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-5, June 2013

[13]Fan Chen, HongJie He, HongXia Wang, “A Fragile Watermarking Scheme for Audio Detection and Recovery”, 978-0-7695-3119-9/08 $25.00 © 2008 IEEE