ALGORITHMS FOR AUDIO WATERMARKING AND STEGANOGRAPHY

NEDELJKO CVEJIC

Department of Electrical and
Information Engineering,
Information Processing Laboratory,
University of Oulu

OULU 2004

NEDELJKO CVEJIC

ALGORITHMS FOR AUDIO WATERMARKING AND STEGANOGRAPHY

Academic Dissertation to be presented with the assent of the Faculty of Technology, University of Oulu, for public discussion in Kuusamonsali (Auditorium YB210), Linnanmaa, on June 29th, 2004, at 12 noon.

OULUN YLIOPISTO, OULU 2004

Supervised by
Professor Tapio Seppänen

Reviewed by
Professor Aarne Mämmelä
Professor Min Wu

ISBN 951-42-7383-4 (nid.)
ISBN 951-42-7384-2 (PDF) http://herkules.oulu.fi/isbn9514273842/

ISSN 0355-3213 http://herkules.oulu.fi/issn03553213/

OULU UNIVERSITY PRESS
OULU 2004

Cvejic, Nedeljko, Algorithms for audio watermarking and steganography
Department of Electrical and Information Engineering, Information Processing Laboratory, University of Oulu, P.O.Box 4500, FIN-90014 University of Oulu, Finland
2004
Oulu, Finland

Abstract

Broadband communication networks and multimedia data available in a digital format opened many challenges and opportunities for innovation. Versatile and simple-to-use software and decreasing prices of digital devices have made it possible for consumers from all around the world to create and exchange multimedia data. Broadband Internet connections and near error-free transmission of data facilitate people to distribute large multimedia files and make identical digital copies of them. A perfect reproduction in digital domain have 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 and steganography 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. In this thesis, we address three research problems on audio watermarking: First, what is the highest watermark bit rate obtainable, under the perceptual transparency constraint, and how to approach the limit? Second, how can the detection performance of a watermarking system be improved using algorithms based on communications models for that system? Third, how can overall robustness to attacks to a watermark system be increased using attack characterization at the embedding side? An approach that combined theoretical consideration and experimental validation, including digital signal processing, psychoacoustic modeling and communications theory, is used in developing algorithms for audio watermarking and steganography.

The main results of this study are the development of novel audio watermarking algorithms, with the state-of-the-art performance and an acceptable increase in computational complexity. The algorithms' performance is validated in the presence of the standard watermarking attacks. The main technical solutions include algorithms for embedding high data rate watermarks into the host audio signal, using channel models derived from communications theory for watermark transmission and the detection and modeling of attacks using attack characterization procedure. The thesis also includes a thorough review of the state-of-the-art literature in the digital audio watermarking.

Keywords: audio watermarking, digital rights management, information hiding, steganography

To my family

Preface

The research related to this thesis has been carried out at the MediaTeam Oulu Group (MT) and the Information Processing Laboratory (IPL), University of Oulu, Finland. I joined the MediaTeam in December 2000 and started my postgraduate studies, leading to the thesis, at the Department of Electrical and Information Engineering in April 2001. Professor Jaakko Sauvola, the director of the MT, docent Timo Ojala, the associate director of the MT, and professor Tapio Seppänen, the MT's scientific director are acknowledged for creating an inspiring research environment of the MT.

I was fortunate to have professor Tapio Seppänen, who was at the time the head of the IPL, as my thesis supervisor. His pursuit for the uppermost standards in research was the great source of my motivation. I wish to thank him for his guidance and encouragement, especially during the starting period of my postgraduate study.

I am grateful to the reviewers of the thesis, professor Min Wu from the University of Maryland, College Park, USA, and professor Aarne Mämmelä from the Technical Research Centre of Finland (VTT), Oulu, Finland. Their feedback improved the quality of the thesis significantly. I am also thankful to Lic. Phil. Pertti Väyrynen for proofreading the manuscript.

I am thankful to my project managers and team leaders Jani Korhonen, Anja Keskinarkaus and Mikko Löytynoja for knowing how to distribute my workload related to the projects and let me carry out research and study that was not always in the narrow scope of the project. I would like to especially thank to Timo Ojala for his credence and support throughout these years. He invested a lot of time and patience in solving numerous practical problems and in making my life in Oulu more pleasant. He would always find time for my dilemmas and our discussions that ranged from research issues to latest happenings in the Premier League.

My special thanks are due to my friends with whom I spent my spare time in Oulu. My first neighbors Ilijana and Djordje Tujkovic were a great source of support and happiness for me. Ilijana was my closest friend that had enough patience to help with all the issues emerging from my immature personality. Djordje, being himself a researcher, was not only a friend to me; he also gave me many advices that had a positive impact to the length of my PhD studies. Anita and Dejan Danilovic, although working hard 12 hours a day, would always find some extra time to hang out with me. I thank them for all the great late night hours we spend together, their sincere friendship and enormous moral

support throughout my studies. The largest part of this thesis was made using the PC that I borrowed from them. Dejan Drajic and Zoran Vukcevic, besides being my friends, had a specific role of familiarizing Finland to me and giving me advices that helped me a lot in the everyday life. Dejan Drajic and Jonne Miettunen were my favorite pub mates and "football experts" that I liked to argue with. I thank Sharat Khungar for all the late lunches we had together in Aularavintola and all the new things I learned about the culture of the Indian subcontinent.

I wish also to thank to Protic family, my first cousins Nemanja and Aleksandar and my aunt Jelena and uncle Zivadin. Thank you for your love and support, not only during my PhD studies, but also throughout the hard times my family went trough.

The financial support provided by Infotech Oulu Graduate School, Nokia, Sonera, Yomi, the National Technology Agency of Finland (TEKES), the Nokia Foundation, and the Tauno Tönning Foundation is gratefully acknowledged.

It is hard find words to express my gratitude to my loving parents, Bogdanka and Slavko for everything they have done for me. Thank you for your love, guidance, as well as encouragement that you have unquestioningly given to me. I thank sincerely to my brother Dejan for standing by my side during all ups and downs in my life, for his immense support, love and credence. My dedication to hard work and vigor to face all the good and less pleasant things that life brings, I grasp from your love and support you have given to me.

Oulu, May 2004 Nedeljko Cvejic

List of Contributions

This thesis is based on the ten original papers (Appendices I–X) which are referred in the text by Roman numerals. All analysis and simulation results presented in publications or this thesis have been produced solely by the author. Professor Tapio Seppänen gave guidance and needed expertise in general signal processing methods. He had an important role in the development of the initial ideas and shaping of the final outline of the publications.

I Cvejic N, Keskinarkaus A & Seppänen T (2001) Audio watermarking using m sequences and temporal masking. In Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, NY, October 2001, p. 227–230.

II Cvejic N & Seppänen T (2001) Improving audio watermarking performance with HAS-based shaping of pseudo-noise. In Proc. IEEE International Symposium on Signal Processing and Information Technology, Cairo, Egypt, December 2001, p. 163–168.

III Cvejic N & Seppänen T (2002) Audio prewhitening based on polynomial filtering for optimal watermark detection. In Proc. European Signal Processing Conference, Toulouse, France, September 2002, p. 69–72.

IV Cvejic N & Seppänen T (2002) A wavelet domain LSB insertion algorithm for high capacity audio steganography. In Proc. IEEE Digital Signal Processing Workshop, Callaway Gardens, GA, October 2002, p. 53–55.

V Cvejic N & Seppänen T (2002) Increasing the capacity of LSB-based audio steganography. In Proc. IEEE International Workshop on Multimedia Signal Processing, St. Thomas, VI, December 2002, p. 336–338.

VI Cvejic N & Seppänen T (2003) Audio watermarking using attack characterization. Electronics Letters 13(39): p. 1020–1021.

VII Cvejic N, Tujkovic D & Seppänen T (2003) Increasing capacity of an audio watermark channel using turbo codes. In Proc. IEEE International Conference on Multimedia and Expo (ICME'03), Baltimore, MD, July 2003, p. 217–220.

VIII Cvejic N & Seppänen T (2003) Rayleigh fading channel model versus AWGN channel model in audio watermarking. In Proc. Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, November 2003, p. 1913-1916.

IX Cvejic N & Seppänen T (2004) Spread spectrum audio watermarking using frequency hopping and attack characterization. Signal Processing 84(1): p. 207–213.

X Cvejic N & Seppänen T (2004) Increasing robustness of an improved spread spectrum audio watermarking method using attack characterization. In Proc. International Workshop on Digital Watermarking, Lecture Notes in Computer Science 2939: p. 467–473.

The general spread-spectrum methods used partially in Paper I and for some other publications (see references) were developed in cooperation with M.Sc. Anja Keskinarkaus. The contribution of Dr. Djordje Tujkovic in Paper VII was expertise in the area of fading channels and channel coding. He also provided turbo coding software, crucial for experimental simulations.

Symbols and Abbreviations

A/D	Analog to Digital
AAC	Advanced Audio Coding
AWGN	Additive White Gaussian Noise
BEP	Bit Error Probability
BER	Bit Error Rate
bps	Bits Per Second
CD	Compact Disc
CSI	Channel State Information
D/A	Digital to Analog
DC	Direct Current
DFT	Discrete Fourier Transform
DS	Direct Sequence
DSP	Digital Signal Processing
DVD	Digital Versatile Disc
DWT	Discrete Wavelet Transform
FFT	Fast Fourier Transform
FH	Frequency Hopping
FIR	Finite Impulse Response
GTC	Gain of Transform Coding
HAS	Human Auditory System
HVS	Human Visual System
ID	Identity
IID	Independent Identically Distributed
ISS	Improved Spread Spectrum
ISO	International Organization for Standardization
IWT	Integer Wavelet Transform
JND	Just Noticeable Distortion
LSB	Least Significant Bit
MER	Minimum-Error Replacement
MPEG	Moving Picture Experts Group
mp3	MPEG 1 Compression, Layer 3
MSE	Mean-Squared Error

NMR	Noise to Mask Ratio (in decibels)
PDA	Personal Digital Assistant
PDF	Probability Density Function
PN	Pseudo Noise
PRN	Pseudo Random Noise
PSC	Power-Density Spectrum Condition
QIM	Quantization Index Modulation
SDMI	Secure Digital Music Initiative
SMR	Signal to Mask Ratio (in decibels)
SNR	Signal to Noise Ratio (in decibels)
SPL	Sound Pressure Level
SS	Spread Spectrum
SYNC	Synchronization
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
VHS	Video Home System
WMSE	Weighted Mean-Squared Error
WEP	Word Error Probability
WER	Word Error Rate

$\mathbf{A}_{\omega k}$	Fourier Coefficients of the Watermarked Signal
$\mathbf{b}$	Binary Encoded Watermark Message
$\widehat{b}$	Decoded Binary Watermark Message
$\mathbf{c}_o$	Host Signal
c_t^{ij}	Cost Function
$\mathbf{c}_w$	Watermarked Signal
$\mathbf{c}_{wn}$	Received Signal
C	Channel Capacity
C_h	Capacity of L Parallel Channels
C_i	Magnitude of an FFT Coefficient
d_{emb}	Embedding Distortion
d_{att}	Attack Distortion
$\mathbf{f}$	Verification Binary Vector
$\mathbf{G}$	Random Variable That Models the Channel Fading Variation
h	Entropy
I($\mathbf{r}$;m)	Mutual Information Between Transmitted Watermark Message and Received Signal $\mathbf{r}$
$\mathbf{k}$	Key Sequence
K	Watermark Key
L	Number of Parallel Channels in Signal Decomposition
L_b	Length of Vector $\mathbf{b}$
L_x	Length of Vector $\mathbf{x}$
$\mathbf{m}$	Watermark Message
m	Subband Index
$\mathbf{n}$	Random Noise

$N_{Re,Im}(\omega)$	Integer Quantized Value
$\mathbf{o}[f]$	Observation Sequence
p_{f_n}	False Negative Probability
p_{f_p}	False Positive Probability
p_x(x)	L_x-dimensional Probability Density Function
Q	Normalized Correlation
Q(**r**;**s**)	Probability Matrix
r	Received Signal
r	Sufficient Statistics at Receiver
$\mathbf{R}^+$	Set of all Positive Real Numbers
R	Redundancy Factor in Spread Spectrum Communications
R	Coding Gain
s	Watermarked Signal
S	Pooled Sample Standard Error
S_i	Quantization Step Size
T_0^2, T_1^2	Test Statistics
T_i	Audibility Threshold
v(t)	Fading Parameter
w	Watermark Sequence
$\mathbf{w}_a$	Added Pattern
$\mathbf{w}_n$	Noisy Added Pattern
$\mathbf{w}_{ri}$	Reference Pattern
W_x^L(K)	Codebook Encrypted in the Watermark Key K
x	Host Signal
$\mathbf{Z}_g$	Gaussian Distributed Variable
α	Parameter in the Improved Spread Spectrum Scheme
λ	Parameter in the Improved Spread Spectrum Scheme
λ_{opt}	Optimal Parameter λ for the Improved Spread Spectrum Scheme
$\mu(x, b)$	Improved Spread Spectrum Function
θ_n	Weight for the Expected Squared Error Introduced by the nth Data Element
σ^2	Variance of the Quantization Noise
$\phi(z)$	Phase of Audio Signal
$\Phi_k(z)$	Total Phase Modulation
$\psi(t)$	Haar Wavelet

Contents

1 Introduction

The rapid development of the Internet and the digital information revolution caused significant changes in the global society, ranging from the influence on the world economy to the way people nowadays communicate. Broadband communication networks and multimedia data available in a digital format (images, audio, video) opened many challenges and opportunities for innovation. Versatile and simple-to-use software and decreasing prices of digital devices (e.g. digital photo cameras, camcorders, portable CD and mp3 players, DVD players, CD and DVD recorders, laptops, PDAs) have made it possible for consumers from all over the world to create, edit and exchange multimedia data. Broadband Internet connections and almost an errorless transmission of data facilitate people to distribute large multimedia files and make identical digital copies of them.

Digital media files do not suffer from any quality loss due to multiple copying processes, such as analogue audio and VHS tapes. Furthermore, recording medium and distribution networks for analogue multimedia are more expensive. These first-view advantages of digital media over the analogue ones transform to disadvantages with respect to the intellectual rights management because a possibility for unlimited copying without a loss of fidelity cause a considerable financial loss for copyright holders [1, 2, 3]. The ease of content modification and a perfect reproduction in digital domain have promoted the protection of intellectual ownership and the prevention of the unauthorized tampering of multimedia data to become an important technological and research issue [4].

A fair use of multimedia data combined with a fast delivery of multimedia to users having different devices with a fixed quality of service is becoming a challenging and important topic. Traditional methods for copyright protection of multimedia data are no longer sufficient. Hardware-based copy protection systems have already been easily circumvented for analogue media. Hacking of digital media systems is even easier due to the availability of general multimedia processing platforms, e.g. a personal computer. Simple protection mechanisms that were based on the information embedded into header bits of the digital file are useless because header information can easily be removed by a simple change of data format, which does not affect the fidelity of media.

Encryption of digital multimedia prevents access to the multimedia content to an individual without a proper decryption key. Therefore, content providers get paid for the delivery of perceivable multimedia, and each client that has paid the royalties must be able to decrypt a received file properly. Once the multimedia has been decrypted, it can

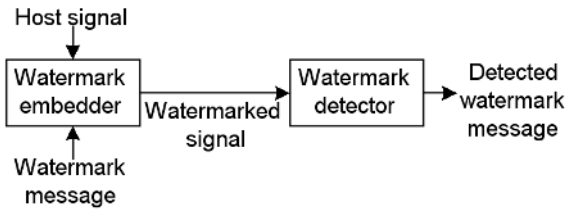

Fig. 1.1. **A block diagram of the encoder.**

be repeatedly copied and distributed without any obstacles. Modern software and broadband Internet provide the tools to perform it quickly and without much effort and deep technical knowledge. One of the more recent examples is the hack of the Content Scrambling System for DVDs [5, 6]. It is clear that existing security protocols for electronic commerce serve to secure only the communication channel between the content provider and the user and are useless if commodity in transactions is digitally represented.

Digital watermarking has been proposed as a new, alternative method to enforce the intellectual property rights and protect digital media from tampering. It involves a process of embedding into a host signal a perceptually transparent digital signature, carrying a message about the **host signal** in order to "mark" its ownership. The digital signature is called the **digital watermark**. The digital watermark contains data that can be used in various applications, including digital rights management, broadcast monitoring and tamper proofing. Although perceptually transparent, the existence of the watermark is indicated when watermarked media is passed through an appropriate watermark detector.

Figure 1.1 gives an overview of the general watermarking system [2]. A watermark, which usually consists of a binary data sequence, is inserted into the host signal in the **watermark embedder**. Thus, a watermark embedder has two inputs; one is the watermark message (usually accompanied by a secret key) and the other is the host signal (e.g. image, video clip, audio sequence etc.). The output of the watermark embedder is the **watermarked signal**, which cannot be perceptually discriminated from the host signal. The watermarked signal is then usually recorded or broadcasted and later presented to the **watermark detector**. The detector determines whether the watermark is present in the tested multimedia signal, and if so, what message is encoded in it. The research area of watermarking is closely related to the fields of information hiding [7, 8] and steganography [9, 10]. The three fields have a considerable overlap and many common technical solutions. However, there are some fundamental philosophical differences that influence the requirements and therefore the design of a particular technical solution. **Information hiding** (or **data hiding**) is a more general area, encompassing a wider range of problems than the watermarking [2]. The term **hiding** refers to the process of making the information imperceptible or keeping the existence of the information secret. **Steganography** is a word derived from the ancient Greek words *steganos* [2], which means *covered* and

graphia, which in turn means *writing*. It is an art of concealed communication.

Therefore, we can define **watermarking systems** as systems in which the hidden message is related to the host signal and **non-watermarking** systems in which the message is unrelated to the host signal. On the other hand, systems for embedding messages into host signals can be divided into **steganographic systems**, in which the existence of the message is kept secret, and **non-steganographic systems**, in which the presence of the embedded message does not have to be secret. Division of the information hiding systems into four categories is given in Table 1.1 [2].

	Host Signal Dependent Message	**Host Signal Independent Message**
Message Hidden	Covert Communication	Steganographic Watermarking
Message Known	Non-steganographic Watermarking	Overt Embedded Communications

Table 1.1. Four categories of information hiding systems.

The primary focus of this thesis is the watermarking of digital audio (i.e., **audio watermarking**), including the development of new watermarking algorithms and new insights of effective design strategies for audio steganography. The watermarking algorithms were primarily developed for digital images and video sequences [11, 12]; interest and research in audio watermarking started slightly later [13, 14]. In the past few years, several algorithms for the embedding and extraction of watermarks in audio sequences have been presented. All of the developed algorithms take advantage of the perceptual properties of the human auditory system (HAS) in order to add a watermark into a host signal in a perceptually transparent manner. Embedding additional information into audio sequences is a more tedious task than that of images, due to dynamic supremacy of the HAS over human visual system [11]. In addition, the amount of data that can be embedded transparently into an audio sequence is considerably lower than the amount of data that can be hidden in video sequences as an audio signal has a dimension less than two-dimensional video files. On the other hand, many attacks that are malicious against image watermarking algorithms (e.g. geometrical distortions, spatial scaling, etc.) cannot be implemented against audio watermarking schemes.

1.1 Scope of research

1.1.1 Application areas

Digital watermarking is considered as an imperceptible, robust and secure communication of data related to the host signal, which includes embedding into and extraction from the host signal. The basic goal is that embedded watermark information follows the watermarked multimedia and endures unintentional modifications and intentional removal attempts. The principal design challenge is to embed watermark so that it is reliably

detected in a watermark detector. The relative importance of the mentioned properties significantly depends on the application for which the algorithm is designed. For copy protection applications, the watermark must be recoverable even when the watermarked signal undergoes a considerable level of distortion, while for tamper assessment applications, the watermark must effectively characterize the modification that took place. In this section, several application areas for digital watermarking will be presented and advantages of digital watermarking over standard technologies examined.

Ownership Protection

In the **ownership protection** applications, a watermark containing ownership information is embedded to the multimedia host signal. The watermark, known only to the copyright holder, is expected to be very robust and secure (i.e., to survive common signal processing modifications and intentional attacks), enabling the owner to demonstrate the presence of this watermark in case of dispute to demonstrate his ownership. Watermark detection must have a very small false alarm probability. On the other hand, ownership protection applications require a small embedding capacity of the system, because the number of bits that can be embedded and extracted with a small probability of error does not have to be large.

Proof of ownership

It is even more demanding to use watermarks not only in the identification of the copyright ownership, but as an actual **proof of ownership**. The problem arises when adversary uses editing software to replace the original copyright notice with his own one and then claims to own the copyright himself. In the case of early watermark systems, the problem was that the watermark detector was readily available to adversaries. As elaborated in [2], anybody that can detect a watermark can probably remove it as well. Therefore, because an adversary can easily obtain a detector, he can remove owner's watermark and replace it with his own. To achieve the level of the security necessary for proof the of ownership, it is indispensable to restrict the availability of the detector. When an adversary does not have the detector, the removal of a watermark can be made extremely difficult. However, even if owner's watermark cannot be removed, an adversary might try to undermine the owner. As described in [2], an adversary, using his own watermarking system, might be able to make it appear as if his watermark data was present in the owner's original host signal. This problem can be solved using a slight alteration of the problem statement. Instead of a direct proof of ownership by embedding e.g. "Dave owns this image" watermark signature in the host image, algorithm will instead try to prove that the adversary's image is derived from the original watermarked image. Such an algorithm provides indirect evidence that it is more probable that the real owner owns the disputed image, because he is the one who has the version from which the other two were created.

Authentication and tampering detection

In the content authentication applications, a set of secondary data is embedded in the host multimedia signal and is later used to determine whether the host signal was tampered. The robustness against removing the watermark or making it undetectable is not a concern as there is no such motivation from attacker's point of view. However, forging a valid authentication watermark in an unauthorized or tampered host signal must be

prevented. In practical applications it is also desirable to locate (in time or spatial dimension) and to discriminate the unintentional modifications (e.g. distortions incurred due to moderate MPEG compression [15, 16]) from content tampering itself. In general, the watermark embedding capacity has to be high to satisfy the need for more additional data than in ownership protection applications. The detection must be performed without the original host signal because either the original is unavailable or its integrity has yet to be established. This kind of watermark detection is usually called a **blind detection**.

Fingerprinting

Additional data embedded by watermark in the fingerprinting applications are used to trace the originator or recipients of a particular copy of multimedia file [17, 18, 19, 20, 21, 22, 23, 24, 25]. For example, watermarks carrying different serial or identity (ID) numbers are embedded in different copies of music CDs or DVDs before distributing them to a large number of recipients. The algorithms implemented in fingerprinting applications must show high robustness against intentional attacks and signal processing modifications such as lossy compression or filtering. Fingerprinting also requires good anti-collusion properties of the algorithms, i.e. it is not possible to embed more than one ID number to the host multimedia file, otherwise the detector is not able to distinguish which copy is present. The embedding capacity required by fingerprinting applications is in the range of the capacity needed in copyright protection applications, with a few bits per second.

Broadcast monitoring

A variety of applications for audio watermarking are in the field of broadcasting [26, 27, 28, 29]. Watermarking is an obvious alternative method of coding identification information for an active broadcast monitoring. It has the advantage of being embedded within the multimedia host signal itself rather than exploiting a particular segment of the broadcast signal. Thus, it is compatible with the already installed base of broadcast equipment, including digital and analogue communication channels. The primary drawback is that embedding process is more complex than a simple placing data into file headers. There is also a concern, especially on the part of content creators, that the watermark would introduce distortions and degrade the visual or audio quality of multimedia. A number of broadcast monitoring watermark-based applications are already available on commercial basis. These include program type identification, advertising research, broadcast coverage research etc. Users are able to receive a detailed proof of the performance information that allows them to:

1. Verify that the correct program and its associated promos aired as contracted;
2. Track barter advertising within programming;
3. Automatically track multimedia within programs using automated software online.

Copy control and access control

In the copy control application, the embedded watermark represents a certain copy control or access control policy. A watermark detector is usually integrated in a recording or playback system, like in the proposed DVD copy control algorithm [5] or during the development Secure Digital Music Initiative (SDMI) [30]. After a watermark has been detected and content decoded, the copy control or access control policy is enforced by di-

recting particular hardware or software operations such as enabling or disabling the record module. These applications require watermarking algorithms resistant against intentional attacks and signal processing modifications, able to perform a blind watermark detection and capable of embedding a non-trivial number of bits in the host signal.

Information carrier

The embedded watermark in this application is expected to have a high capacity and to be detected and decoded using a blind detection algorithm. While the robustness against intentional attack is not required, a certain degree of robustness against common processing like MPEG compression may be desired. A public watermark embedded into the host multimedia might be used as the link to external databases that contain certain additional information about the multimedia file itself, such as copyright information and licensing conditions. One interesting application is the transmission of metadata along with multimedia. Metadata embedded in, e.g. audio clip, may carry information about composer, soloist, genre of music, etc.

1.1.2 Research areas

Watermarking algorithms can be characterized by a number of defining properties [2]. Six of them, which are most important for audio watermarking algorithms [31], represent our research subareas. The relative importance of a particular subarea is application-dependent, and in many cases the interpretation of a watermark property itself varies with the application.

Perceptual transparency

In most of the applications, the watermark-embedding algorithm has to insert additional data without affecting the perceptual quality of the audio host signal [11, 32]. The fidelity of the watermarking algorithm is usually defined as a perceptual similarity between the original and watermarked audio sequence. However, the quality of the watermarked audio is usually degraded, either intentionally by an adversary or unintentionally in the transmission process, before a person perceives it. In that case, it is more adequate to define the fidelity of a watermarking algorithm as a perceptual similarity between the watermarked audio and the original host audio at the point at which they are presented to a consumer.

Watermark bit rate

The bit rate of the embedded watermark is the number of the embedded bits within a unit of time and is usually given in bits per second (bps). Some audio watermarking applications, such as copy control, require the insertion of a serial number or author ID, with the average bit rate of up to 0.5 bps. For a broadcast monitoring watermark, the bit rate is higher, caused by the necessity of the embedding of an ID signature of a commercial within the first second at the start of the broadcast clip, with an average bit rate up to 15 bps. In some envisioned applications, e.g. hiding speech in audio or compressed audio stream in audio, algorithms have to be able to embed watermarks with the bit rate that is

a significant fraction of the host audio bit rate, up to 150 kbps.

Robustness

The robustness of the algorithm is defined as an ability of the watermark detector to extract the embedded watermark after common signal processing manipulations. A detailed overview of robustness tests is given in Chapter 3. Applications usually require robustness in the presence of a predefined set of signal processing modifications, so that watermark can be reliably extracted at the detection side. For example, in radio broadcast monitoring, embedded watermark need only to survive distortions caused by the transmission process, including dynamic compression and low pass filtering, because the watermark detection is done directly from the broadcast signal. On the other hand, in some algorithms robustness is completely undesirable and those algorithms are labeled **fragile audio watermarking** algorithms.

Blind or informed watermark detection

In some applications, a detection algorithm may use the original host audio to extract watermark from the watermarked audio sequence (**informed detection**). It often significantly improves the detector performance, in that the original audio can be subtracted from the watermarked copy, resulting in the watermark sequence alone. However, if detection algorithm does not have access to the original audio (**blind detection**) and this inability substantially decreases the amount of data that can be hidden in the host signal. The complete process of embedding and extracting of the watermark is modeled as a communications channel where watermark is distorted due to the presence of strong interference and channel effects [33]. A strong interference is caused by the presence of the host audio, and channel effects correspond to signal processing operations.

Security

Watermark algorithm must be secure in the sense that an adversary must not be able to detect the presence of embedded data, let alone remove the embedded data. The security of watermark process is interpreted in the same way as the security of encryption techniques and it cannot be broken unless the authorized user has access to a secret key that controls watermark embedding. An unauthorized user should be unable to extract the data in a reasonable amount of time even if he knows that the host signal contains a watermark and is familiar with the exact watermark embedding algorithm. Security requirements vary with application and the most stringent are in cover communications applications, and, in some cases, data is encrypted prior to embedding into host audio.

Computational complexity and cost

The implementation of an audio watermarking system is a tedious task, and it depends on the business application involved. The principal issue from the technical point of view is the computational complexity of embedding and detection algorithms and the number of embedders and detectors used in the system. For example, in broadcast monitoring, embedding and detection must be done in real time, while in copyright protection applications, time is not a crucial factor for a practical implementation. One of the economic issues is the design of embedders and detectors, which can be implemented as hardware or software plug-ins, is the difference in processing power of different devices (laptop,

PDA, mobile phone, etc.).

1.2 Problem statement

1.2.1 Research problem

The fundamental process in each watermarking system can be modeled as a form of communication where a message is transmitted from watermark embedder to the watermark receiver [2]. The process of watermarking is viewed as a transmission channel through which the watermark message is being sent, with the host signal being a part of that channel. In Figure 1.2, a general mapping of a watermarking system into a communications model is given (more details are provided in Chapter 4). After the watermark is embed-

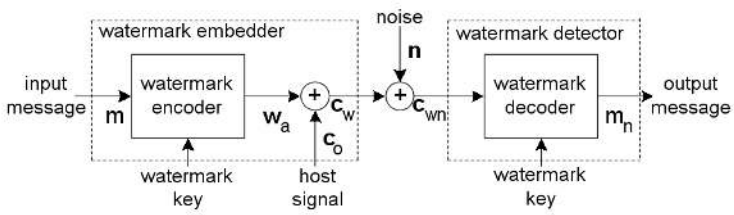

Fig. 1.2. **A watermarking system and an equivalent communications model.**

ded, the watermarked work is usually distorted after watermark attacks. The distortions of the watermarked signal are, similarly to the data communications model, modeled as additive noise.

When setting down the research plan for this study, the research of digital audio watermarking was in its early development stage; the first algorithms dealing specifically with audio were presented in 1996 [11]. Although there were a few papers published at the time, a basic theory foundations were laid down and the concept of the "magic triangle" introduced (Chapter 3). Therefore, it is natural to place watermarking into the framework of the traditional communications system. The main line of reasoning of the "magic triangle" concept (Chapter 3) is that if the perceptual transparency parameter is fixed, the design of a watermark system cannot obtain high robustness and watermark data rate at the same time. Thus, we decided to divide the research problem into three specific subproblems. They are:

SP1: What is the highest watermark bit rate obtainable, under perceptual transparency constraint, and how to approach the limit?

SP2: How can the detection performance of a watermarking system be improved using algorithms based on communications models for that system?

SP3: How can overall robustness to attacks of a watermark system be increased using an attack characterization at the embedding side?

1.2.2 Research hypothesis

The division of the research problem into the three subproblems above define the following three research hypotheses:

RH1: To obtain a distinctively high watermark data rate, embedding algorithm can be implemented in a transform domain, with the usage of the least significant bit coding.

RH2: To improve detection performance, a spread spectrum method can be used, cross correlation between the watermark sequence and host audio decreased and channel coding introduced.

RH3: To achieve the robustness of watermarking algorithms, an attack characterization can be introduced at the embedder, improved channel model can be derived and informed detection can be used for watermark decoding.

1.2.3 Research assumptions

The general research assumption is that the process of embedding and extraction of watermarks can be modeled as a communication system, where the watermark embedding is modeled as a transmitter, the distortion of watermarked signal as a communications channel noise and watermark extraction as a communications detector.

It is also assumed that modeling of the human auditory system and the determination of perceptual thresholds can be done accurately using models from audio coding, namely MPEG compression HAS model [15, 16].

The perceptual transparency (inaudibility) of a proposed audio watermarking scheme can be confirmed through subjective listening tests in a predefined laboratory environment with a participation of a predefined number of people with a different music education and background.

A central assumption in the security analysis of the proposed algorithms is that an adversary that attempts to disrupt the communication of watermark bits or remove the watermark does not have access to the original host audio signal.

1.2.4 Research methods

In this thesis, a multidisciplinary approach is applied for solving the research subproblems. The signal processing methods are used for watermark embedding and extracting processes, derivation of perceptual thresholds, transforms of signals to different signal

domains (e.g. Fourier domain, wavelet domain), filtering and spectral analysis. Communication principles and models are used for channel noise modeling, different ways of signalling the watermark (e.g. a direct sequence spread spectrum method, frequency hopping method), derivation of optimized detection method (e.g. matched filtering) and evaluation of overall detection performance of the algorithm (bit error rate, normalized correlation value at detection). The basic information theory principles are used for the calculation of the perceptual entropy of an audio sequence, channel capacity limits of a watermark channel and during design of an optimal channel coding method. The research methods also include algorithm simulations with real data (music sequences) and subjective listening tests.

1.3 Outline of the thesis

Robust digital audio watermarking algorithms and high capacity steganography methods for audio are studied in this thesis. The purpose of the thesis is to develop novel audio watermarking algorithms providing a performance enhancement over the other state-of-the-art algorithms with an acceptable increase in complexity and to validate their performance in the presence of the standard watermarking attacks. Presented as a collection of ten original publications enclosed as appendices I-X, the thesis is organized as follows.

Chapter 2 introduces the basic concepts and definitions of digital watermarking, in order to place in context the main contributions of the thesis developed as the combination of digital signal processing, psychoacoustic modeling and communications theory. The properties of the HAS that are exploited in the process of audio watermarking are shortly reviewed. A survey of the key digital audio watermarking algorithms is presented subsequently.

A general background and requirements for high capacity covert communications for audio are presented in Chapter 3. A perceptual entropy measure for audio signals and information theoretic assessment of the achievable data rates of a data hiding channel are reviewed. In addition, the results which are in part documented in Papers IV and V, for the modified time domain LSB steganography algorithm and a high bit rate algorithm in wavelet domain are presented.

In Chapter 4, the contents of which are in part included in Papers I, II, III, and VII, several spread spectrum audio watermarking algorithms in time domain are presented. A general model for the spread spectrum-based watermarking is described in order to place in context the developed algorithms. The parts of communication theory, which were used in order to find a relationship between the capacity of the watermarked channel and the distortion caused by a malicious attack, are given in this chapter as well.

Chapter 5, the contents of which are in part presented in Papers VI, VIII, IX, and X, focuses on the increasing of the robustness of embedded watermarks using attack characterization. Novel principles important for our attack characterization implementation are presented, as well as watermark channel models of interest. A method for introducing the attack characterization approach in an improved spread spectrum scheme is discussed.

Chapter 6 concludes the thesis discussing its main results and contributions. Directions for further development and open problems for future research are also described.

2 Literature survey

This chapter reviews the appropriate background literature and describes the concept of information hiding in audio sequences. Scientific publications included into the literature survey have been chosen in order to build a sufficient background that would help out in solving the research subproblems problems stated in Chapter 1. In addition, Chapter 2 presents general concepts and definitions used and developed in more details in Chapters 3, 4 and 5. We decided to divide the theoretical background into three parts, presented in Chapters 3, 4 and 5 because of the specific structure of the thesis, which presents three different concepts for data hiding in audio, contrary to the usual concept of elaborating a single idea. Therefore, the theoretical background in subjunction to the particular concept is given as a separate subchapter in the respective chapters. In this manner, it much easier for the reader to follow the presented concepts, and the chapters themselves can also be read as standalone readings.

In the first section, the properties of the **human auditory system** (HAS) that are exploited in the process of audio watermarking are shortly reviewed. A survey of the key digital audio watermarking algorithms and techniques is presented subsequently. The algorithms are classified by the signal domain in which the watermark is inserted (time domain, Fourier domain, etc.) and statistical method used for the embedding and extraction of watermark bits.

Audio watermarking initially started as a sub-discipline of digital signal processing, focusing mainly on convenient signal processing techniques to embed additional information to audio sequences. This included the investigation of a suitable transform domain for watermark embedding and schemes for the imperceptible modification of the host audio. Only recently watermarking has been placed to a stronger theoretical foundation, becoming a more mature discipline with a proper base in both communication modeling and information theory. Therefore, short overviews of the basics of information theory and channel modeling for watermarking systems are given in this chapter.

2.1 Overview of the properties of the HAS

Watermarking of audio signals is more challenging compared to the watermarking of images or video sequences, due to wider dynamic range of the HAS in comparison with human visual system (HVS) [11]. The HAS perceives sounds over a range of power greater than 10^9:1 and a range of frequencies greater than 10^3:1. The sensitivity of the HAS to the additive white Gaussian noise (AWGN) is high as well; this noise in a sound file can be detected as low as 70 dB below ambient level.

On the other hand, opposite to its large dynamic range, HAS contains a fairly small differential range, i.e. loud sounds generally tend to mask out weaker sounds. Additionally, HAS is insensitive to a constant relative phase shift in a stationary audio signal and some spectral distortions interprets as natural, perceptually non-annoying ones. [11].

Auditory perception is based on the critical band analysis in the inner ear where a frequency-to-location transformation takes place along the basilar membrane. The power spectra of the received sounds are not represented on a linear frequency scale but on limited frequency bands called **critical bands** [34]. The auditory system is usually modeled as a bandpass filterbank, consisting of strongly overlapping bandpass filters with bandwidths around 100 Hz for bands with a central frequency below 500 Hz and up to 5000 Hz for bands placed at high frequencies. If the highest frequency is limited to 24000 Hz, 26 critical bands have to be taken into account.

Two properties of the HAS dominantly used in watermarking algorithms are **frequency (simultaneous) masking** (Section 2.1.1) and **temporal masking** (Section 2.1.2)[34]. The concept using the perceptual holes of the HAS is taken from wideband audio coding (e.g. MPEG compression 1, layer 3, usually called mp3)[16]. In the compression algorithms, the holes are used in order to decrease the amount of the bits needed to encode audio signal, without causing a perceptual distortion to the coded audio. On the other hand, in the information hiding scenarios, masking properties are used to embed additional bits into an existing bit stream, again without generating audible noise in the audio sequence used for data hiding.

2.1.1 Frequency masking

Frequency (simultaneous) masking is a frequency domain phenomenon where a low level signal, e.g. a pure tone (the maskee), can be made inaudible (masked) by a simultaneously appearing stronger signal (the masker), e.g. a narrow band noise, if the masker and maskee are close enough to each other in frequency [34]. A masking threshold can be derived below which any signal will not be audible. The masking threshold depends on the masker and on the characteristics of the masker and maskee (narrowband noise or pure tone). For example, with the masking threshold for the sound pressure level (SPL) equal to 60 dB, the masker in Figure 2.1 at around 1 kHz, the SPL of the maskee can be surprisingly high - it will be masked as long as its SPL is below the masking threshold. The slope of the masking threshold is steeper toward lower frequencies; in other words, higher frequencies tend to be more easily masked than lower frequencies. It should be pointed out that the distance between masking level and masking threshold is smaller in noise-masks-

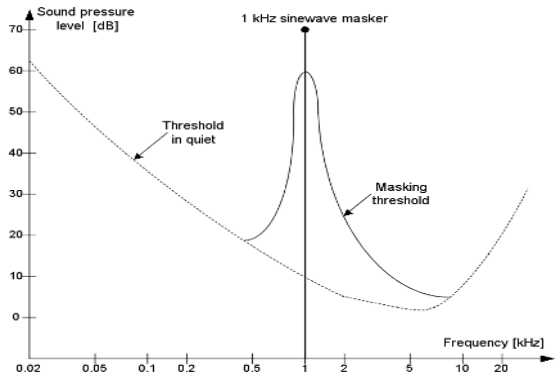

Fig. 2.1. **Frequency masking in the human auditory system (HAS), reference sound pressure level is $\mathbf{p}_0 = 2 \cdot 10^{-5}$ Pa.**

tone experiments than in tone-masks-noise experiments due to HAS's sensitivity toward additive noise. Noise and low-level signal components are masked inside and outside the particular critical band if their SPL is below the masking threshold. Noise contributions can be coding noise, inserted watermark sequence, aliasing distortions, etc. Without a masker, a signal is inaudible if its SPL is below the threshold in quiet, which depends on frequency and covers a dynamic range of more than 70 dB as depicted in the lower curve of Figure 2.1. The qualitative sketch of Figure 2.2 gives more details about the masking threshold. The distance between the level of the masker (given as a tone in Figure 2.2) and the masking threshold is called **signal-to-mask ratio** (SMR) [16]. Its maximum value is at the left border of the critical band. Within a critical band, noise caused by watermark embedding will be audible as long as **signal-to-noise ratio** (SNR) for the critical band [16] is higher than its SMR. Let SNR(m) be the signal-to-noise ratio resulting from watermark insertion in the critical band m; the perceivable distortion in a given subband is then measured by the noise to mask ratio:

$$\text{NMR(m)=SMR-SNR(m)} \tag{2.1}$$

The **noise-to-mask ratio** NMR(m) expresses the difference between the watermark noise in a given critical band and the level where a distortion may just become audible; its value in dB should be negative.

This description is the case of masking by only one masker. If the source signal consists of many simultaneous maskers, a global masking threshold can be computed that describes the threshold of just noticeable distortion (JND) as a function of frequency [34]. The calculation of the global masking threshold is based on the high resolution short-term amplitude spectrum of the audio signal, sufficient for critical band-based analysis and is usually performed using 1024 samples in FFT domain. In a first step, all the individual

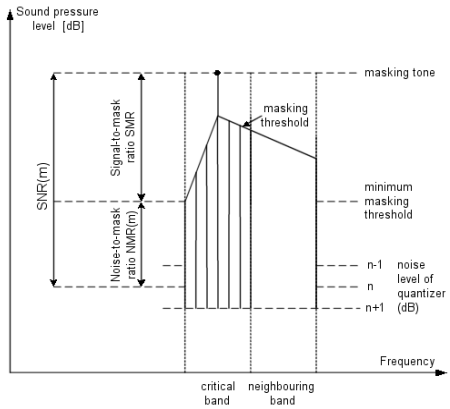

Fig. 2.2. **Signal-to-mask-ratio and Signal-to-noise-ratio values.**

masking thresholds are determined, depending on the signal level, type of masker (tone or noise) and frequency range. After that, the global masking threshold is determined by adding all individual masking thresholds and the threshold in quiet. The effects of the masking reaching over the limits of a critical band must be included in the calculation as well. Finally, the global signal-to-noise ratio is determined as the ratio of the maximum of the signal power and the global masking threshold [16], as depicted in Figure 2.1.

2.1.2 Temporal masking

In addition to frequency masking, two phenomena of the HAS in the time domain also play an important role in human auditory perception. Those are pre-masking and post-masking in time [34]. The temporal masking effects appear before and after a masking signal has been switched on and off, respectively (Figure 2.3). The duration of the pre-masking is significantly less than one-tenth that of the post-masking, which is in the interval of 50 to 200 milliseconds. Both pre- and post-masking have been exploited in the MPEG audio compression algorithm and several audio watermarking methods.

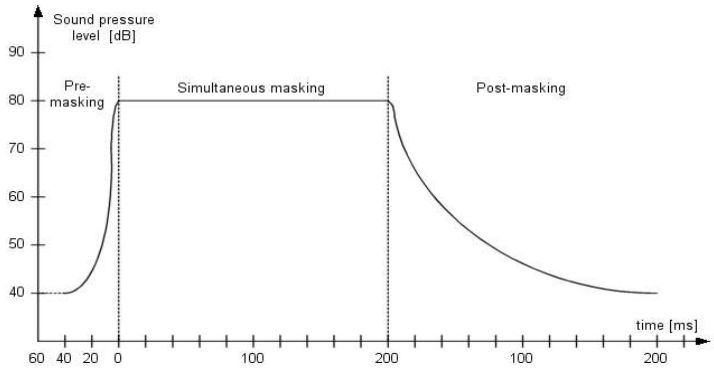

Fig. 2.3. **Temporal masking in the human auditory system (HAS).**

2.2 General concept of watermarking

2.2.1 A general model of digital watermarking

Figure 2.4 gives an overview of the general model of the digital watermarking considered in this chapter. A watermark message **m** is embedded into the host signal **x** to produce the watermarked signal **s**. The embedding process is dependent on the key K and must satisfy the perceptual transparency requirement, i.e. the subjective quality difference between **x** and **s** (denoted as embedding distortion d_{emb}) must be below the just noticeable difference threshold. Before the watermark detection and decoding process takes place, **s** is usually intentionally or unintentionally modified. The intentional modifications are usually referred to as attacks; an attack produce attack distortion d_{att} at a perceptually acceptable level. After attacks, a watermark extractor receives attacked signal **r**.

The watermark extraction process consists of two sub-processes, first, watermark decoding of a received watermark message $\hat{\mathrm{m}}$ using key K, and, second, watermark detection, meaning the hypothesis test between:

Hypothesis H_0: the received data **r** is not watermarked with key K, and
Hypothesis H_1: the received data **r** is watermarked with key K.

Depending on a watermarking application, the detector performs an informed or blind watermark detection. The term **attack** requires some further clarification. Watermarked signal **s** can be modified without the intention to impact the embedded watermark (e.g. dynamic amplitude compression of audio prior to radio broadcasting). Why is this kind of signal processing is called an attack? The first reason is to simplify the notation of the general model of digital watermarking. The other, an even more significant reason, is

that any common signal processing impairing an embedded watermark drastically will be a potential method applied by adversaries that intentionally try to remove the embedded watermark. The watermarking algorithms must be designed to endure the worst possible attacks for a given attack distortion d_{att}, which might be even some common signal processing operation (e.g. dynamic compression, low pass filtering etc.). Furthermore, it is generally assumed that the adversary has only one watermarked version **s** of the host signal **x**. In fingerprinting applications, differently watermarked data copies could be exploited by collusion attacks. It has been proven that robustness against collusion attacks can be achieved by a sophisticated coding of different watermark messages embedded into each data copy [23]. However, it seems that the necessary codeword length increases dramatically with the number of watermarked copies available to the adversary.

The separation between watermark decoding and watermark detection during the watermark extraction should be clearly defined as well. Thus, it is important to differ between communicating a watermark message **m** (embedding and decoding of a digital watermark) and verifying whether the received data **r** is watermarked or not (watermark detection). At first glance, the decision between the hypotheses H_0 and H_1 (watermark detection) appears as a special case of decoding a binary watermark message $m \in \{0, 1\}$. This is not the case because in binary watermark communication the watermarked signal and received signal have some special composition for m=0 and another special structure for m=1. However, in the hypothesis H_0 of the detection problem, the received data can have any structure or, equivalently, no structure at all.

The importance of the key K has to be emphasized. The embedded watermarks should be secure against detection, decoding, removal or modification or modification by adversaries. Kerckhoff's principle [35], stating that the security of a crypto system has to reside only in the key of a system, has to be applied when the security of a watermarking system is analyzed. Therefore, it must be assumed that the watermark embedding and extraction algorithms are publicly known, but only those parties knowing the proper key are able to receive and modify the embedded information. The key K is considered a large integer number, with a word length of 64 bits to 1024 bits. Usually, a key sequence **k** is derived from K by a cryptographically secure random number generator to enable a secure watermark embedding for each element of the host signal.

Several more detailed models of watermarking systems, including modeling of water-

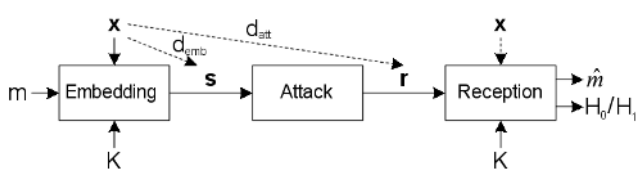

Fig. 2.4. **General model of digital watermarking.**

mark channel with encryption, are given in Chapter 4. Since three communication theory based audio watermarking algorithms are described in Chapter 4, we decided to place more detailed overview of the modeling the watermarking systems using data communications models in there, including all the relevant references.

2.2.2 Statistical modeling of digital watermarking

In order to properly analyze digital watermarking systems, a stochastic description of the multimedia data is required. The watermarking of data whose content is perfectly known to the adversary is useless. Any alteration of the host signal could be inverted perfectly, resulting in a trivial watermarking removal. Thus, essential requirements on data being robustly watermarkable are that there is enough randomness in the structure of the original data and that quality assessments can be made only in a statistical sense. In this section, basic statistical modeling of digital watermarking is introduced and general assumptions are explained.

Let the original host signal **x** be a vector of length L_x. Statistical modeling of data means to consider **x** a realization of a discrete random process $\boldsymbol{x}$ [6]. In the most general form, $\boldsymbol{x}$ is described by an L_x-dimensional probability density function (PDF) $p_x(x)$.

$$p_x(\boldsymbol{x}) = \prod_{n=1}^{L_x} p_{x_n}(x_n) \tag{2.2}$$

with $p_{x_n}(x_n)$ being the nth marginal distribution of x. A further simplification is to assume independent, identically distributed (IID) data elements so $p_{x_n}(x_n) = p_{x_j}(x_n) = p_x(x)$. Most multimedia data cannot be modeled adequately by an IID random process [6]. However, in many cases, it is possible to decompose the data into components such that each component can be considered almost statistically independent. In most cases, the multimedia data have to be transformed, or parts have to be extracted, to obtain a component-wise representation with mutually independent and IID components. The watermarking of independent data components can be considered as communication over parallel channels.

Watermarking embedding and attacks against digital watermarks must be such that the introduced perceptual distortion - the subjective difference between the watermarked and attacked signal to the original host signal is acceptable. In the previous section, we introduced the terms embedding distortion d_{emb} and attack distortion d_{att}, but no specific definition was given. The definition of an appropriate objective distortion measure is crucial for the design and analysis of a digital watermarking system. A useful objective distortion measure must be convenient for the statistical analysis of watermarking use-cases and should be appropriate for the quality evaluation of real-world multimedia data.

The weighted mean-squared error (WMSE) distortion measure is adopted in the published work in the field, as it usually offers a good compromise between appropriateness for multimedia signals and convenience for statistical analysis. For a WMSE distortion

measure, the embedding distortion d_{emb} and attack distortion d_{att} are given by [6]

$$d_{emb} = D(\mathbf{x}, \mathbf{s}, \Theta) = \frac{1}{L_x} \sum_{n=1}^{L_x} \Theta_n E\left\{(x_n - s_n)^2\right\}, \tag{2.3}$$

$$d_{att} = D(\mathbf{x}, \mathbf{r}, \Theta) = \frac{1}{L_x} \sum_{n=1}^{L_x} \Theta_n E\left\{(x_n - r_n)^2\right\}. \tag{2.4}$$

In (2.3) and (2.4) E$\{\cdot\}$ denotes expectation and $\Theta_n \in \mathbf{R}^+$ is the weight for the expected squared error introduced in the nth data element. x_n, s_n, and r_n are the nth elements of the host audio $\mathbf{x}$, watermarked sequence $\mathbf{s}$ and received signal $\mathbf{r}$, respectively. The weight Θ_n lets a simple adaptation of the objective distortion measure to the subjectively different importance of data elements. For IID data, the weights Θ_n are usually set to 1 since none of the data elements is subjectively preferred and the WMSE is reduced to the simple mean-squared error (MSE) distortion measure [6]. Furthermore, the WMSE distortion measure fits well to the component-wise data description introduced above. It is very common that identical weights Θ_j can be used for all elements of the jth data component. For example, the weighted embedding distortion in the discrete wavelet domain (DWT) [36] can be written as

$$d_{emb} = D\left(\left\{\mathbf{x}_j^{DWT}\right\}, \left\{\mathbf{s}_j^{DWT}\right\}, \left\{\Theta_j^{DWT}\right\}\right) = \frac{1}{J} \sum_{j=1}^{J} \Theta_j^{DWT} E\left\{\left(x_j^{DWT} - s_j^{DWT}\right)^2\right\} \tag{2.5}$$

where $\mathbf{x}_j^{DWT}$ represents the jth element of the host audio sequence $\mathbf{x}$ in wavelet domain and $\mathbf{s}_j^{DWT}$ stands for the jth element of the watermarked sequence $\mathbf{s}$ in wavelet domain. In practice, an adversary can never evaluate d_{emb} since he does not know $\mathbf{x}$. On the other hand, it is fair to assume, during watermark embedding, that an adversary could obtain a good approximation of d_{emb}. In contrast, measuring the attack distortion at the detection side, by $D(\mathbf{s}, \mathbf{r}, \Theta)$, which is practical for an adversary, might be misleading since a perfect attack ($\mathbf{r} = \mathbf{x}, D(\mathbf{s}, \mathbf{x}, \Theta) > 0$) would be rated worse than no attack ($\mathbf{r} = \mathbf{s}, D(\mathbf{s}, \mathbf{s}, \Theta) = 0$).

The performance of different watermarking schemes for specific stochastic data is extensively analyzed in the literature. It is usually assumed that the embedder and the attacker have access to the same stochastic model. Within this framework, provable limits for optimal watermarking schemes and optimal attacks can be derived. In practice, provable limits are difficult to obtain, because an improvement of the available statistical models for the data at hand can help an adversary as well.

2.2.3 Decoding and detection performance evaluation

The ultimate goal of any watermarking algorithm is a reliable watermark extraction. In general, extraction reliability for a specific watermarking scheme relies on the features of the original data, on the embedding distortion d_{emb} and on the attack distortion d_{att}. Watermark extraction reliability is usually analyzed for different levels of attack distortion

d_{att} and fixed data features and embedding distortion d_{emb}. Different reliability measures are used for watermark decoding and watermark detection.

2.2.3.1 Watermark decoding

In the performance evaluation of the watermark decoding, digital watermarking is considered as a communication problem. A watermark message **m** is embedded into the host signal **x** and must be reliably decodable from the received signal **r** [6]. Low decoding error rates can be achieved only using error correction codes. For practical error correcting coding scenarios, the watermark message is usually encoded into a vector **b** of length L_b with binary elements $b_n = 0, 1$. Usually, **b** is also called the binary watermark message, and the decoded binary watermark message is $\hat{\mathbf{b}}$. The decoding reliability of **b** can be described by the word error probability (WEP)

$$p_w = P_r(m \neq \hat{m}) = P_r(\mathbf{b} \neq \hat{\mathbf{b}}), \tag{2.6}$$

or by the bit error probability (BEP)

$$p_b = \frac{1}{L_b} \sum_{n=1}^{l_b} P_r(b_n \neq \hat{b}_n). \tag{2.7}$$

The WEP and BEP can be computed for specific stochastic models of the entire watermarking process including attacks. The predicted error probabilities can be confirmed experimentally by a large number of simulations with different realizations of the watermark key K, the host signal **x**, the attack parameters and a watermark message **m**. The number of measured error events divided by the number of the observed events defines the measured error rates, word error rate, WER and bit error rate BER.

Performance limits can be derived with methods borrowed from the information theory. For example, the maximum watermark rate which can be received in principle without errors is determined by the mutual information $I(\mathbf{r}|m)$ between the transmitted watermark message m and received data **r** and given by [37]

$$I(\mathbf{r}|m) = h(\mathbf{r}) - h(\mathbf{r}|m) \tag{2.8}$$

where $h(\mathbf{r})$ is the differential entropy of **r** and $h(\mathbf{r}|m)$ is the differential entropy of **r** conditioned on the transmitted message **m**. The PDFs $p_r(\mathbf{r})$ and $p_r(r|m = \mathbf{m})$ are required for the computation of $h(\mathbf{r})$ and $h(\mathbf{r}|m)$. $I(\mathbf{r}|m)$ can be achieved only for an infinite number of data elements. For a finite number of data elements, a non-zero word error probability p_w or a bit error probability p_b are unavoidable.

The channel capacity C of a specific communication channel is defined as the maximum mutual information $I(r; m)$ over all transmissions schemes with a transmission power constrained to a fixed value [37]. The watermark capacity C is defined correspondingly with a slight modification specific for watermarking scenarios. The capacity analysis provides a good method for comparing the performance limits of different communication scenarios, and thus is frequently employed in the existing literature. Since

there is still no solution available for the general watermarking problem, digital watermarking is usually analyzed within certain constraints on the embedding and attacks. Additionally, for different scenarios, the watermark capacity might depend on different parameters (domain of embedding, attack parameters, etc.).

2.2.3.2 Watermark detection

Watermark detection is defined as the decision whether the received data **r** is watermarked (H_1) or not watermarked (H_0) [6]. In general, both hypotheses cannot be separated perfectly. Thus, we define the probability p_{fp} (false positive) as the case of accepting H_1 when H_0 is true and the probability p_{fn} of accepting H_0 when H_1 is true (false negative). In many applications, the hypothesis test must be designed to ensure a limited false positive probability, e.g. $p_{fp} < 10^{-12}$ for watermark detection in the context of DVD copy protection. Another option for the evaluation of watermark detection is the investigation of the total detection error probability p_e, which measures both possible error types.

In this thesis, the watermark detection is based on watermarking schemes that have been designed for reliable communication of a binary watermark message **b**. A subvector **f** of length L_f of the watermark message **b** is used for a validity verification of a received watermark message $\hat{\mathbf{b}}$. Without a loss of generality, an all-zero verification message can be used since the security of the embedded watermark is ensured by a key sequence **k** derived from the key K. Two simple watermark detection methods using the verification bit vector **f** are discussed. In the first method, a detection based on a hard decision decoded verification is applied. In the second method, known encoded verification bits are exploited to implement detection based on so-called soft values, where the soft values are obtained by a further processing of the received signal.

Hard decision decoding

The verification message **f** is encoded together with all remaining watermark message bits to obtain the encoded watermark message $\mathbf{b}_c$. During the watermark extraction, the message $\hat{\mathbf{b}}$ is as in the communication scenario. One fraction of $\hat{\mathbf{b}}$ is the decoded watermark verification message $\hat{\mathbf{f}}$ that must be equal to **f** for a valid watermark message $\hat{\mathbf{b}}$. Therefore, the hypothesis decision rule is given by:

$$H_0 : \hat{\mathbf{f}} \neq \mathbf{f} \tag{2.9}$$

$$H_1 : \hat{\mathbf{f}} = \mathbf{f} \tag{2.10}$$

The false positive probability p_{fp} can be calculated based on the assumption that $P_r(\hat{f}_n = 0|H_0) = P_r(\hat{f}_n = 1|H_1) = 0.5$. The probability $p_{fp} = 0.5_{L_f}$ is obtained for L_f independent bits $\hat{f}_n$ and depends only on the number L_f of verification bits. The false negative probability depends on the bit error probability p_b and the number of verification bits

$$p_{fn} = 1 - (1 - p_b)^{L_f} \,. \tag{2.11}$$

In the expression (2.11) statistically independent received verification bits $\hat{f}_n$ are assumed. In practice, the interleaving of all bits in **b** before error correction encoding is useful to

ensure the validity of those assumptions. A generalization of the decision rule given above is to accept H_1 if the Hamming distance [37], $d_H(\hat{\mathbf{f}_n}, \mathbf{f})$ is lower than a certain threshold. In that case, the threshold could be designed to find a better trade-off between p_{fp} and p_{fn}.

Soft decision decoding

Detection based on a hard decision decoding is very simple. However, if the accurate statistical models of the introduced attacks are known, soft decision decoding gives potentially a better detection performance. The verification message $\mathbf{f}$ is equal to the first L_f bits of $\mathbf{b}$ and error correction coding of $\mathbf{b}$ is such that the first L_{f_c} bits of the coded watermark message $\mathbf{b}_c$ are independent of the remaining watermark message bits. Without a loss of generality, we can assume

$$(b_{c,0}, ..., b_{c,L_{f_c}-1}) = \mathbf{f} = \mathbf{0}. \tag{2.12}$$

Let I_f denote the set of the indices of all data elements with embedded coded verification bits. We assume that the PDFs $P_r(r_{I_f}|H_0)$ and $P_r(r_{I_f}|H_1)$ for receiving r_{I_f} depending on hypothesis H_0 or H_1, respectively, are known. Bayes' solution to the hypothesis testing problem can be applied, which is

$$\frac{P_r(r_{I_f}|H_1)}{P_r(r_{I_f}|H_0)} > T \Rightarrow \text{accept H}_1, \text{else} \Rightarrow \text{accept H}_0 \tag{2.13}$$

where T is the decision threshold. T is a constant depending on the a priori probabilities for H_1 and H_0 and the cost connected with different decision errors. For $T = 1$, the decision rule above forms a maximum-likelihood (ML) detector. For equal a priori probabilities, the decision error probability is $p_e = \frac{1}{2}(p_{fp} + p_{fn})$. Assuming equal a priori probabilities and equal costs for both hypotheses, the above decision rule can be reformulated so that H_1 is accepted if

$$Pr = \frac{P_r(r_{I_f}|H_1)}{P_r(r_{I_f}|H_1) + P_r(r_{I_f}|H_0)} > 0.5 \tag{2.14}$$

where $Pr \in [0, 1]$ denotes the reliability that a received watermark message $\hat{\mathbf{b}}$ is a valid watermark message. For decision above, p_{fn} and p_{fp} depend directly on the PDFs $P_r(r_{I_f}|H_0)$ and $P_r(r_{I_f}|H_1)$.

2.2.4 Exploiting side information during watermark embedding

In most blind watermarking schemes, as in a blind spread spectrum watermarking, the host signal is considered as interfering noise during the watermark extraction. Nevertheless, recently it has been realized that a blind watermarking can be modeled as communication with side information at the encoder. This has been published in [38] and [39] independently. The main idea is that, although the blind receiver does not have access to the host signal $\mathbf{x}$, the encoder can exploit knowledge of $\mathbf{x}$ to reduce the influence of the host signal on the watermark detection and decoding. In [39], general concepts based on an

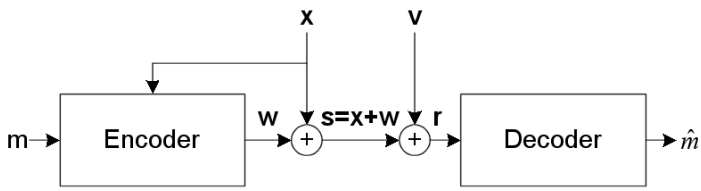

Fig. 2.5. Communication with side information at the encoder over an AWGN channel.

early paper by Shannon [40] are described. Therein, the usefulness of side information at the encoder is shown, without any detailed data of the principal improvements or the optimal exploitation of the side information. Also, one of the assumptions of Shannon's paper is that the encoder knows only a causal part of the host signal **x**. Chen and Wornell introduced a paper by Costa [41] from year 1983 to the watermarking community. Costa considered communication with side information at the encoder over an AWGN channel as depicted in Figure 2.5. A scheme that was derived by him performs well as if the original data (the side information at the encoder) were perfectly known to the decoder. Chen and Wornell showed that their previously developed watermarking scheme [38] based on quantization index modulation (QIM) can be considered as a part of Costa's scheme and that an extended version of QIM can perform as well as Costa's scheme. It is purely theoretical and thus several practical approaches to implement Costa's scheme were proposed [42, 43]. Figure 2.5 depicts a block diagram of the considered watermarking embedding into IID host signal **x** of length L_x and blind detection. The watermark message $m \in 1, 2, \ldots, M$ is embedded with a constraint on the embedding distortion d_{emb}. The embedding process exploiting side information of the host signal is separated into two parts: first, an appropriate watermark sequence **w** representing the watermark message **m** is selected, and, second, **w** is added to the host signal **x**. The MSE distortion measure is used so that

$$d_{emb} = \frac{1}{L_x} E\left\{\|\mathbf{s} - \mathbf{x}\|^2\right\} = \frac{1}{L_x} E\left\{\|\mathbf{w}\|^2\right\}. \tag{2.15}$$

The mapping of **m** onto sequence **w**, also of length L_x, is determined by **x** and the by codebook $W^{L_x}(K)$, which is encrypted with the watermark key K. Secrecy is obtained by a pseudo-random selection of all entries in $W^{L_x}(K)$.

The assumption is that the watermark sequences **w** are zero mean and IID. The embedding distortion d_{emb} is then equal to the variance σ_w^2 of the watermark elements w_n. The AWGN attack is independent of the characteristics of the original and watermarked signal so that attack distortion is $d_{att} = d_{emb} + \sigma_v^2 = \sigma_w^2 + \sigma_v^2$. It should be noted that a blind spread spectrum watermarking (Section 2.3.4) also fits into the given model. For the spread spectrum watermarking, the codebook $W^{L_x}(K)$ contains all combinations of possible messages **m** and of spreading sequences derived from K, which is a finite number of sequences. Furthermore, the performance limit of an optimal non-blind watermarking

scheme can also be considered as the ultimate performance limit of blind watermarking.

2.2.5 The information theoretical approach to digital watermarking

Early research on watermarking can be characterized by an alternating advancement of watermarking schemes and attacks. A theoretical approach to digital watermarking should give answers about the convergence of this process. Some work in this direction has been published independently by Su et al. [44, 45] and by Moulin et al. [46, 47]. In [44], a power-density spectrum condition (PSC) for watermark signal has been derived, which ensures that a linear estimation of embedded watermark is as hard as possible. Independently, Moulin et al. [46] introduced the notion of the "information hiding game". Information theoretic and game-theoretic concepts are exploited to set up a well-defined theoretical framework for digital watermarking. In [46], Moulin et al. discuss the case of watermarking IID Gaussian host signal. Extensions of this work to non-white Gaussian original data and application to image watermarking have been developed by Su et al. [44, 45] and Moulin et al. [46, 47].

A conceptual description of a watermarking game is given in [46]. Assume watermarking of the host signal **x** with some statistical properties is investigated. First, a nonnegative distortion function for the host signal **x** of length L_x is defined. Second, the watermarking process has to be characterized. This contains:
• The set of watermark messages M
• The embedding function depending on the watermark message **m** and key K and constrained to the embedding distortion d_{emb} and
• The decoding function, which depends on the key K.
Third, the attack channel, constrained by the attack distortion d_{att}, is defined by the probability matrix $Q(\mathbf{r}, \mathbf{s})$ describing the mapping of a certain watermarked signal **s** of length L_x to a certain attacked signal **r** of length L_x in a statistical sense.

A watermarking process with embedding distortion d_{emb} and attack channel with an attack distortion d_{att} define the watermarking game between the embedded and attacker subject to distortion pair (d_{emb}, d_{att}). One suitable objective function of the game is the achievable watermark rate. A certain watermark rate is achievable for (d_{emb}, d_{att}) if there is a watermark process subject to embedding distortion d_{emb} with rates $R' > R$ such that the probability of the decoding error goes to zero as the signal length L_x goes to infinity, for any attack channel subject to attack distortion d_{att}.

The watermark capacity $C(d_{emb}, d_{att})$ is the supremum of all achievable rates R for distortions d_{emb}, d_{att}. The watermark capacity is achieved if the embedder chooses a watermarking process that maximizes the achievable rate R while the attacker chooses an attack channel that minimizes the achievable rate R. A compete solution to the above described general watermarking game is currently not available. Thus, suboptimal watermarking schemes, e.g. SS watermarking and suboptimal attack channels, for example, AWGN attacks, are considered.

2.3 Selected audio watermarking algorithms

Watermarking algorithms were primarily developed for digital images and video sequences; interest and research in audio watermarking started slightly later. In the past few years, several algorithms for the embedding and extraction of watermarks in audio sequences have been presented. All of the developed algorithms take advantage of the perceptual properties of the human auditory system (HAS) in order to add a watermark into a host signal in a perceptually transparent manner. A broad range of embedding techniques goes from simple **least significant bit** (LSB) scheme to the various spread spectrum methods. The overview given in this section presents the best known general audio watermarking algorithms, with an emphasis on the algorithms that were used as a basis for published work (LSB algorithm, spread spectrum, improved spread spectrum, etc).

In the notation used throughout the section, $x[i], i = 1, \ldots, l(c_o)$ are the samples of the host audio signal in the time domain. The range of the values of the audio signal is $x[i] \in [-1, 1)$, with 16-bit amplitude resolution, providing $2^{16} = 65536$ quantization levels in total. An additional index of the host audio sequence $\mathbf{x}_{oj}$ denotes a subset of the host audio. As a large majority of the audio watermarking algorithms use various overlapping and nonoverlapping blocks in order to embed data, $x_j[i]$ is used to represent the ith sample in the jth block of size $l(x_j)$. Individual blocks of the host audio are used to embed part of one bit, one bit, a number of bits or a complete watermark $\mathbf{m}$.

2.3.1 LSB coding

One of the earliest techniques studied in the information hiding and watermarking area of digital audio (as well as other media types [48, 49, 50]) is LSB coding [51, 52]. A natural approach in the case of the audio sequences is to embed watermark data by alternation of the individual samples of the digital audio stream having the amplitude resolution of 16 bits per sample. It usually does not use any psychoacoustics model to perceptually weight the noise introduced by LSB replacement. However, as will be elaborated in the Chapter 3, we developed a novel method to introduce a certain level of perceptual shaping of the LSB coding.

The watermark encoder uses a subset of all available host audio samples $\mathbf{x}$ chosen by a secret key. The substitution operation $x_j[i] \rightarrow m[i]$ on the LSBs is performed on this subset. The extraction process simply retrieves the watermark by reading the value of these bits. Therefore, the decoder needs all the samples of the watermarked audio that were used during the embedding process. Usually, $l(\mathbf{x}_o) \gg l(\mathbf{m})$. Thus the robustness of the method can be improved by a repeated watermark embedding. The modification of the LSBs of the samples used for data hiding introduces a low power additive white Gaussian noise. As noted in the previous Chapter, HAS is very sensitive to the AWGN and this fact limits the number of LSBs that can be imperceptibly modified.

The main advantage of the method is a very high watermark channel capacity; the use of only one LSB of the host audio sample gives capacity of 44.1 kbps. The obvious disadvantage is the extremely low robustness of the method, due to fact that random changes of the LSBs destroy the coded watermark [53]. In addition, it is very unlikely that embedded

watermark would survive digital to analogue and subsequent analogue to digital conversion. Since no computationally demanding transformation of the host signal in the basic version of this method needs to be done, this algorithm has a very small algorithmic delay. This permits the use on this LSB in real-time applications. This algorithm is a good basis for steganographic applications for audio signals and a base for steganalysis [54, 55, 56].

2.3.2 Watermarking the phase of the host signal

Algorithms that embed watermark into the phase of the host audio signal do not use masking properties of the HAS, but the fact that the HAS is insensitive to a constant relative phase shift in a stationary audio signal [11]. There are two main approaches used in the watermarking of the host signal's phase, first, phase coding [11, 57] and, second, phase modulation [58, 59, 60].

The basic **phase coding** method was presented in [11]. The basic idea is to split the original audio stream into blocks and embed the whole watermark data sequence into the phase spectrum of the first block. One drawback of the phase coding method is a considerably low payload because only the first block is used for watermark embedding. In addition, the watermark is not dispersed over the entire data set available, but is implicitly localized and can thus be removed easily by the cropping attack. It is a non-blind watermarking method (as the phase modulation algorithm) that limits the number of applications it is suitable for.

The watermark insertion in the **phase modulation** method is performed using an independent multiband phase modulation [61, 62]. Imperceptible phase modifications are exploited in this approach by the controlled phase alternation of the host audio. To ensure perceptual transparency by introducing only small changes in the envelope, the performed phase modulation has to satisfy the following constraint

$$|\Delta\phi(z)/\Delta z| < 30^o, \tag{2.16}$$

where $\phi(z)$ denotes the signal phase and z is the **Bark** scale. Each Bark constitutes one critical bandwidth; the conversion of frequency between Bark and Hz is given in [31]. Using a long block size N (e.g. $N = 2^{14}$) algorithm attains a slow phase change over time. The watermark is converted into a phase modulation by having one integer Bark scale carry one message bit of the watermark, with the frequency in Hz. The robustness of the modulated phase can be increased by using multiple Bark values carrying one watermark bit.

The watermark extraction requires a perfect synchronization procedure to perform a block alignment for each watermarked block, using the original signal as a reference. A matching of the particular segments of the modulated phase to the encoded watermark bits is possible if no significant distortions of the watermarked signal took place.

The data rate of the watermark depends on three factors: first, the amount of the redundancy added, second, the frequency range used for watermark embedding, and, third, the energy distribution of the host audio. If the selected Bark's energy is too low, that Bark should be skipped during the watermark embedding procedure. For audio signals sampled at 44.1 kHz, 0-15 kHz (0-24 in Bark scale) proved to be a sensible range for watermark

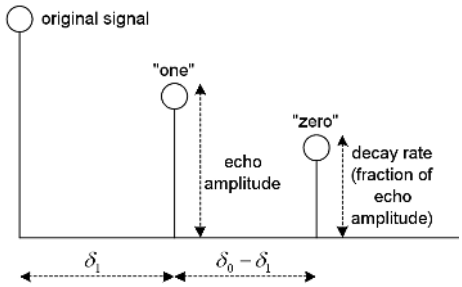

Fig. 2.6. **Parameters of echo embedding watermarking method.**

embedding. If, for example, two Barks carry one watermark bit, the watermark data rate is $(24/2)(44100/2^{14}) = 32$bps.

2.3.3 Echo hiding

A number of developed audio watermarking algorithms [63, 64, 65] are based on echo hiding method, described for the first time in [11]. Echo hiding schemes embed watermarks into a host signal by adding echoes to produce watermarked signal. The nature of the echo is to add resonance to the host audio. Therefore the acute problem of sensitivity of the HAS towards the additive noise is circumvented in this method. After the echo has been added, watermarked signal retains the same statistical and perceptual characteristics.

The offset (or delay) between the original and a watermarked signal is small enough that the echo is perceived by the HAS as an added resonance. The four major parameters, the initial amplitude, decay rate, "one" offset and "zero" offset are given in Figure 2.6. The watermark embedding process can be represented as a system that has one of two possible system functions. In the time domain, the system functions are discrete time exponentials, differing only in the delay between impulses. Processing host signal through any kernel in Figure 2.6 will result in an encoded signal. The delay (number of sample intervals) between the original signal and the echo is dependent on the kernel being used, 1 if the "one" kernel is used and 0 if the "zero" kernel is used.

The host signal is divided into smaller portions for encoding more than one bit. Each individual portion can then be considered as an independent signal and echoed with the desired bit. The final watermarked signal (containing several bits) is a composite of all independently encoded signal portions. A smooth transition between portions encoded with different bits should be adjusted using different methods to prevent abrupt changes in the resonance in the watermarked signal. Information is embedded into a signal by echoing the original signal with one of two delay kernels. Therefore, the extraction of the

embedded information is to detect the spacing between the echoes. The magnitude of the autocorrelation of the encoded signal's cepstrum

$$F^{-1}\left\{\log\left(|F(x)|^2\right)\right\} \tag{2.17}$$

where F represents the Fourier Transform and F^{-1} the inverse Fourier Transform can be examined at two locations, corresponding to the delays of the "one" and "zero" kernel, respectively. If the autocepstrum is greater at δ_1 than it is at δ_0, an embedded bit is decoded as "one". For the multiple echo hiding, all peaks present in the autocepstrum are detected. The number of the peaks corresponding to the delay locations of the "one" and "zero" kernels are then counted and compared. If there are more peaks at the delay locations for the "one" echo kernel, the watermark bit is decoded as "one".

Increased robustness of the watermark algorithm requires high-energy echoes to be embedded which increases audible distortion. There are several modifications to the basic echo-hiding algorithm. Xu et al. [66] proposed a multi-echo embedding technique to reduce the possibility of echo detection by third parties. The technique has clear constraints regarding the increase of the robustness, because the audio timbre is noticeably changed with the sum of pulse amplitude [67]. Oh et al. [67] proposed an echo kernel comprising multiple echoes by both positive and negative pulses with different offsets (closely located) in the kernel, of which the frequency response is smooth in lower bands and has large ripples in high frequency. Although these large ripples are perceptually less significant for a large majority of audio sequences, they can become audible as an unpleasant noise in the sections where audio signal contains low energy.

2.3.4 Spread spectrum watermarking

In a number of the developed algorithms [68, 69, 70, 71, 72], the watermark embedding and extraction are carried out using spread-spectrum (SS) technique. SS sequence can be added to the host audio samples in time domain [68, 73, 74], to FFT coefficients [72, 75, 76, 77], in subband domain [14, 78, 79, 80, 81], to cepstral coefficients [82, 83] and in a compressed domain [84, 85]. If embedding takes place in a transform domain, it should be located in the coefficients invariant to common watermark attacks as amplitude compression, resampling, lowpass filtering, and other common signal processing techniques. The idea is that after the transform, any significant change in the signal would significantly decrease the subjective quality of the watermarked audio.

Watermark is spread over a large number of coefficients and distortion is kept below the just noticeable difference level (JND) by using the occurrence of masking effects of the human auditory system (HAS). Change in each coefficient can be small enough to be imperceptible because the correlator detector output still has a high signal to noise ratio (SNR), since it despreads the energy present in a large number of coefficients. A general model for SS-based watermarking is shown in Figure 2.7. Vector **x** is considered to be the original host signal already in an appropriate transform domain. The vector **y** is the received vector, in the transform domain, after channel noise. A **secret key K** is used by a **pseudo random number generator** (PRN) [86, 87] to produce a chip sequence with

zero mean and whose elements are equal to $+\sigma_u$ or $-\sigma_u$. The sequence **u** is then added to or subtracted from the signal **x** according to the variable b, where b assumes the values of +1 or -1 according to the bit (or bits) to be transmitted by the watermarking process (in multiplicative algorithms multiplication operation is performed instead addition [88]). The signal **s** is the watermarked audio signal. A simple analysis of SS-based watermarking leads to a simple equation for the probability of error. Thus, we define inner product and norm as [89]: $\langle \mathbf{x}, \mathbf{u} \rangle = \sum_{i=0}^{N-1} x_i u_i$ and $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$ where N is the length of the vectors **x**, **s**, **u**, **n**, and **y** in Figure 2.7. Without a loss of generality, we assume that we are embedding one bit of information in a vector **s** of N transform coefficients. Then, the bit rate is $1/N$ bits/sample. That bit is represented by the variable b, whose value is either +1 or -1. Embedding is performed by

$$\mathbf{s} = \mathbf{x} + b\mathbf{u} \tag{2.18}$$

The distortion in the embedded signal is defined by $\|\mathbf{s} - \mathbf{x}\|$. It is easy to see that for the embedding equation (2.23), we have

$$D = \|b\mathbf{u}\| = \|\mathbf{u}\| = \sigma_u. \tag{2.19}$$

The channel is modeled as an additive noise channel $\mathbf{y} = \mathbf{s} + \mathbf{n}$, and the watermark extraction is usually performed by the calculation of the normalized sufficient statistics r:

$$r = \frac{\langle \mathbf{y}, \mathbf{u} \rangle}{\langle \mathbf{u}, \mathbf{u} \rangle} = \frac{\langle b\mathbf{u} + \mathbf{x} + \mathbf{n}, \mathbf{u} \rangle}{\sigma_u} = b + c_x + c_n \tag{2.20}$$

and estimating the embedded bit as $\hat{b}$ =sign(r), where $c_x = \langle \mathbf{x}, \mathbf{u} \rangle / \|\mathbf{u}\|$ and $c_n = \langle \mathbf{n}, \mathbf{u} \rangle / \|\mathbf{u}\|$. Simple statistical models for the host audio **x** and the attack noise **n** are assumed. Namely, both sequences are modeled as uncorrelated white Gaussian random processes: $x_i \sim N(0, \sigma_x^2)$ and $n_i \sim N(0, \sigma_n^2)$. Then, it is easy to show that the sufficient statistics r are also Gaussian variables, i.e.:

$$r \sim N(m_r, \sigma_r^2), m_r = E[r] = b, \sigma_r^2 = \frac{\sigma_x^2 + \sigma_n^2}{N\sigma_u^2} \tag{2.21}$$

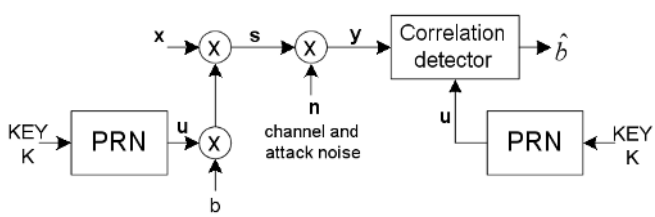

Fig. 2.7. **General model for SS-based watermarking.**

Specifically, let us elaborate the case when b is equal to 1. In that case, an error occurs when $r < 0$, and therefore, the error probability p is given by

$$p = Pr\left\{\hat{b} < 0 | b = 1\right\} = \frac{1}{2}\text{erfc}\left(\frac{m_r}{\sigma_r\sqrt{2}}\right) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\sigma_u^2 N}{2(\sigma_x^2 + \sigma_n^2)}}\right) \tag{2.22}$$

where erfc$(\cdot)$ is complementary error function. The equal error probability is obtained under the assumption that $b = -1$. A plot of that probability as a function of the SNR (in this case defined as (m_r/σ_r) is given in Figure 2.8. For example, from Figure 2.8, it can

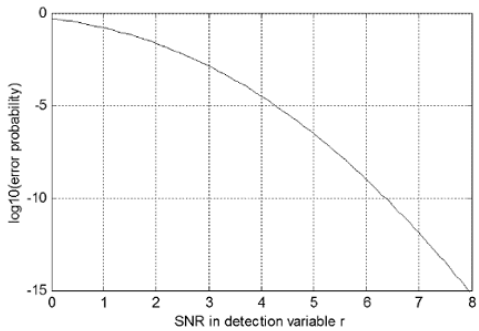

Fig. 2.8. **Error probability as a function of the SNR.**

be seen that if an error probability lower than 10^{-3} is needed, SNR becomes:

$$\frac{m_r}{\sigma_r} > 3 \Rightarrow N\sigma_u^2 > 9\left(\sigma_x^2 + \sigma_n^2\right) \tag{2.23}$$

or more generally, to achieve an error probability p we need:

$$N\sigma_u^2 > 2\left(\text{erfc}^{-1}(p)\right)^2\left(\sigma_x^2 + \sigma_n^2\right) \tag{2.24}$$

Equation (2.29) shows that we can make a trade-off between the length of the chip sequence N with the energy of the sequence σ_u^2. It lets us to simply compute either N or σ_u^2, given the other variables involved.

2.3.5 Improved spread spectrum algorithm

The development of the improved spread spectrum (ISS) method was gradual and consisted of several phases. In [39], the authors described the importance of decreasing the

influence of the host signal on the watermark extraction process, analyzing a spread spectrum system with the fixed cross correlation value. Using framework from [39], in [90] the authors have derived three different watermarking approaches, corresponding to the cases of "maximized robustness", "maximized correlation coefficient" and "constant robustness". Still, the problem of minimizing the bit error rate at a fixed average distortion level during watermark embedding process is not addressed. Final ISS method has been proposed in [91]. It removes the host signal as a source of interference, gaining significantly on the robustness of watermark detection.

The main idea behind the ISS is that by using the encoder knowledge about the signal (or more precisely, c_x, the projection of $\mathbf{x}$ on the watermark), we can enhance performance by modulating the energy of the inserted watermark to compensate for the signal interference. The new embedding approach is defined by a slight modification to the SS embedding, i.e. the amplitude of the inserted chip sequence is varied by a function $\mu(c_x, b)$:

$$\mathbf{s} = \mathbf{x} + \mu(c_x, b)\mathbf{u} \tag{2.25}$$

where, as in the standard SS method, $c_x = \langle \mathbf{x}, \mathbf{u} \rangle / \|\mathbf{u}\|$. It is obvious that the traditional SS is a particular case of ISS. In this notation, SS is a special case of the ISS in which the function μ is made independent of c_x. The simplest version of the ISS is to restrict μ to be a linear function. Not only is this much simpler to analyze, but it also provides a significant part of the gains in relation to traditional SS. In this case, and due to the symmetry of the problem in relation to c_x and b, we have

$$\mathbf{s} = \mathbf{x} + (\alpha b - \lambda c_x)\mathbf{u} \tag{2.26}$$

The parameters α and λ control the distortion level and the removal of the carrier distortion on the detection statistics. Traditional SS is obtained by setting $\alpha = 1$ and $\lambda = 0$. If AWGN channel model is used as we did for SS method, $\mathbf{y} = \mathbf{s} + \mathbf{n}$, the receiver sufficient statistics are:

$$r = \frac{\langle \mathbf{y}, \mathbf{u} \rangle}{\|\mathbf{u}\|} = \alpha b + (1 - \lambda)c_x + c_n \tag{2.27}$$

Therefore, as λ tends to 1, the more the influence of c_x is removed from r. The detector is the same as in SS, i.e., the detected bit is sign(r). The expected distortion of the ISS system is given by:

$$E[D] = E[\|\mathbf{s} - \mathbf{x}\|] = E[\|\alpha b - \lambda c_x\|^2 \sigma_u^2] = \left(\alpha^2 + \frac{\lambda^2 \sigma_x^2}{N\sigma_u^2}\right)\sigma_u^2 \tag{2.28}$$

To force the average distortion of the ISS system to be equal to that of the traditional SS system, we force $E[D] = \sigma_u^2$ and therefore

$$\alpha = \sqrt{\frac{N\sigma_u^2 - \lambda^2 \sigma_x^2}{N\sigma_u^2}} \tag{2.29}$$

In order to compute the error probability, the mean and the variance of the sufficient statistic r are needed. They are given by

$$m_r = \alpha b, \sigma_r^2 = \frac{\sigma_n^2 + (1 - \lambda)^2 \sigma_x^2}{N\sigma_u^2} \tag{2.30}$$

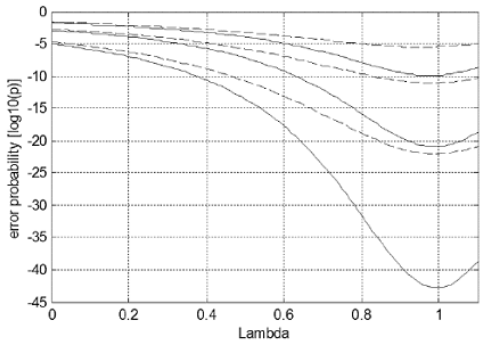

Fig. 2.9. Error probability as a function of λ. Solid lines represent a 10 dB SNR, and dashed lines represent a 7 dB SNR. The three lines correspond to values of equal to N·WNR=5, 10, and 20 (with higher values having smaller error probability).

Thus, the error probability of the ISS system can be computed as:

$$p = Pr\left\{r < 0 | b = 1\right\} = \frac{1}{2}\text{erfc}\left(\frac{m_r}{\sigma_r\sqrt{2}}\right) = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{N\sigma_u^2 - \lambda^2\sigma_x^2}{2(\sigma_n^2 + (1-\lambda)^2\sigma_x^2)}}\right) \quad (2.31)$$

Error probability function p can be rewritten as a function of the watermark-to-noise ratio (WNR) σ_u^2/σ_x^2 and the SNR σ_x^2/σ_n^2 [91]

$$p = \frac{1}{2}\text{erfc}\left(\frac{1}{\sqrt{2}}\sqrt{\frac{\frac{N\sigma_u^2}{\sigma_x^2} - \lambda^2}{\frac{\sigma_n^2}{\sigma_x^2} + (1-\lambda)^2}}\right). \quad (2.32)$$

In Figure 2.9, we plot p as a function of λ for various values of SNR and N·WNR. Note that by a proper selection of the parameter λ, the error probability in the proposed method can be made several orders of magnitude better than using traditional SS. For example, with a signal-to-interference ratio of 10 (i.e., 10 dB), we get a reduction in the error rate from $p_0 = 10^{-5}$ for traditional SS to $p = 1.55 \cdot 10^{-43}$ for the ISS method, which is a reduction of over 37 orders of magnitude in the error probability. Higher SNR values, which can happen in practical applications, lead to even higher gains. As it can be inferred from Figure 2.9, the error probability varies with λ, with the optimum value usually close to one. The expression for the optimum value for can be computed [91] from the error probability by setting $\delta p/\delta\lambda = 0$ and is given by

$$\lambda_{OPT} = \frac{1}{2}\left(\left(1 + \frac{\sigma_n^2}{\sigma_x^2} + \frac{N\sigma_u^2}{\sigma_x^2}\right) - \sqrt{\left(1 + \frac{\sigma_n^2}{\sigma_x^2} + \frac{N\sigma_u^2}{\sigma_x^2}\right)^2 - 4\frac{N\sigma_u^2}{\sigma_x^2}}\right) \quad (2.33)$$

In addition, it is clear from that for N large enough, $\lambda_{OPT} \to 1$ as $SNR \to \infty$.

2.3.6 Methods using patchwork algorithm

The patchwork technique was first presented in [11, 92] for embedding watermarks in images. It is a statistical method based on hypothesis testing and relying on large data sets. As a second of CD quality stereo audio contains 88200 samples, a patchwork approach is applicable for the watermarking of audio sequences as well. The watermark embedding process uses a pseudorandom process to insert a certain statistic into a host audio data set, which is extracted with the help of numerical indexes (like the mean value), describing the specific distribution. The method is usually applied in a transform domain (Fourier, wavelet, etc.) in order to spread the watermark in time domain and to increase robustness against signal processing modifications [93, 94, 95]. Embedding steps are summarized as follows:

1. Map the secret key and the watermark to the seed of a random number generator. After that, generate an index set $I = \{I_1, \ldots, I_{2n}\}$ whose elements are pseudo-randomly selected integer values from $[K_1, K_2]$, where $1 \leq K_1 \leq K_2 \leq N$. Note that two index sets, I^0 and I^1, are needed to denote watermark bits 0 and 1, respectively. The choice of K_1 and K_2 is a crucial step in embedding the watermark because these values control the trade-off between the robustness and the inaudibility of the watermark.

2. Let $F = \{F_1, \ldots, F_N\}$ be the coefficients whose subscript denote frequency range from the lowest to the highest frequencies. Define $A = a_1, \ldots, a_n$ as the subset of F whose subscript corresponds to the first n elements of the index set I^0 or I^1 according to the embedded code with similar definition for $B = b_1, \ldots, b_n$ with the last n elements, that is $a_i = F_I$ and $b_i = F_{In+I}$, for $i = 1, \ldots, n$.

3. Calculate the sample means $\bar{a} = \frac{1}{n}\sum_{i=1}^{n} a_i$ and $\bar{b} = \frac{1}{n}\sum_{i=1}^{n} b_i$, respectively and the pooled sample standard error:

$$S = \sqrt{\frac{\sum_{i=1}^{n}(a_i - \bar{a})^2 + \sum_{i=1}^{n}(b_i - \bar{b})^2}{n(n-1)}} \tag{2.34}$$

4. The embedding function presented below introduces a location-shift change

$$a_i^* = a_i + sign(\bar{a} - \bar{b})\sqrt{C}\frac{S}{2}, b_i^* = b_i - sign(\bar{a} - \bar{b})\sqrt{C}\frac{S}{2} \tag{2.35}$$

5. Finally, replace the selected elements a_i and b_i by a_i^* and b_i^*, respectively, and then apply the inverse DCT.

Since the proposed embedding method introduces relative changes of two sets in location, a natural test statistic which is used to decide whether or not the watermark is embedded should concern the distance between the means of A and B. Thus, the watermark extracting process is done as follows:

1. Map the secret key and watermark to the seed of random number generator and then generate the index sets I^0 and I^1, which was applied to the encoding process.

2. Obtain the subsets A_1 and B_1 from $F = \{F_1, \ldots, F_N\}$ and compute the sample means and the pooled sample standard errors. Obtain the subsets $A_0 = \{a_{01}, , a_{0n}\}$ and $B_0 = \{b_{01}, , b_{0n}\}$ from the index set I^0, $A_1 = \{a_{11}, , a_{1n}\}$ and $B_1 = \{b_{11}, , b_{1n}\}$ from the index set I^1, all from $F = \{F_1, \ldots, F_N\}$ and compute the sample means $\bar{a}_0$, $\bar{a}_1$, $\bar{b}_0$ and $\bar{b}_1$ and the pooled standard errors S_0 and S_1.

3. Calculate the test statistics

$$T_0^2 = \frac{(\bar{a}_0 - \bar{b}_0)^2}{S_0^2}, T_1^2 = \frac{(\bar{a}_1 - \bar{b}_1)^2}{S_1^2} \tag{2.36}$$

and define T^2 as the larger value obtained from two statistics.

4. Compare T^2 with the threshold M and decide that watermark is embedded if $T^2 > M$. Only when $T^2 > M$, bit 0 is assigned if $T_0^2 > T_1^2$, and bit 1 otherwise.
Therefore, the patchwork technique can be observed as the linear comparator function in the spread-spectrum technique.

2.3.7 Methods using various characteristics of the host audio

Several audio watermarking algorithms developed in the recent years use different statistical properties of the host audio and modify them in order to embed watermark data. Those properties are pitch values, number of salient points, difference in energy of two adjacent blocks, etc. However, modifications of the host signal statistical properties do influence the subjective quality of the audio signal and have to be performed in a way that does not produce distortions above the audible threshold. Usually, these methods are robust to signal processing modifications, but offer a low watermark capacity.

Papers [96, 97] introduced content-adaptive segmentation of the host audio according to its characteristics in time domain. Since the embedding parameters are dependent of the host audio, it is along the right direction to increase tamper resistance. The basic idea is to classify the host audio into a predetermined number of segments according to its properties in time domain, and encode each segment with an embedding scheme, which is designed to best suit this segment of audio signal, according to its features in frequency domain.

In paper [98], the temporal envelope of the audio signal is modified according to the watermark. A number of signal processing operations are needed for embedding a multi-bit payload watermark. First, the filter extracts the part of the audio signal that is suitable to carry the watermark information. The watermarked audio signal is then obtained by adding an appropriately scaled version of the product of watermark and filtered host audio to the host signal. Watermark detector consists of two stages: the symbol extraction stage and the correlation and decision stage.

The algorithm presented in [99] embeds the watermark by deciding for each mute period in the host audio whether to extend it by a predefined value. In order to detect the watermark, the detector must have access to the original length of all mute periods in the host audio.

The method described in [100] uses the pitch scaling of the host audio, realized using short time Fourier transform, to embed the watermark. The correlation ratio, computed during the embedding procedure is quantized with different quantization steps in order to embed bit 0 and 1 of the watermark stream.

In papers [101, 102], salient points are used as basis for watermark embedding resistant to desynchronization attacks. A *salient point* is defined as the point in time where the variation of energy of the host audio signal has a large positive peak; it defines the synchronization point for the watermarking process without embedding additional synchronization tags. The embedding of the watermark bits in [101] is performed using a statistical mean manipulation of the cepstral coefficients and in [102] by altering the distance between two salient points.

The algorithms presented in [103, 104] use a feature extraction of the host audio signal in order to tailor the specific embedding algorithm for the given segment of the host audio. In [103], authors use neural networks for the feature extraction and classification, while in [104] the feature extraction is done using a nonlinear frequency scale technique.

2.4 Summary

Chapter 2 reviews the literature and describes the concept of information hiding in audio sequences. Scientific publications included in the literature survey have been chosen in order to build a sufficient background that would help out in solving the research subproblems stated in Chapter 1.

In the first section, the properties of the **human auditory system** (HAS) that are exploited in the process of audio watermarking are shortly reviewed. A survey of the key digital audio watermarking algorithms and techniques is presented subsequently. The algorithms are classified by the signal domain in which the watermark is inserted and statistical method used for embedding and extraction of watermark bits. Audio watermarking initially started as a sub-discipline of digital signal processing, focusing mainly on convenient signal processing techniques to embed additional information to audio sequences. This included the investigation of a suitable transform domain for watermark embedding and schemes for imperceptible modification of the host audio. Only recently has watermarking been placed to a stronger theoretical foundation, becoming a more mature discipline with a proper base in both communication modeling and information theory. Therefore, short overviews of the basics of information theory and channel modeling for watermarking systems were given in this chapter.

3 High capacity covert communications

The simplest visualization of the requirements of information hiding in digital audio is so called **magic triangle** [7], given in Figure 3.1. Inaudibility, robustness to attacks, and the watermark data rate are in the corners of the magic triangle. This model is convenient for a visual representation of the required trade-offs between the capacity of the watermark data and the robustness to certain watermark attacks, while keeping the perceptual quality of the watermarked audio at an acceptable level. It is not possible to attain high robustness to signal modifications and high data rate of the embedded watermark at the same time. Therefore, if a high robustness is required from the watermarking algorithm, the bit rate of the embedded watermark will be low and vice versa, high bit rate watermarks are usually very fragile in the presence of signal modifications. However, there are some applications that do not require that the embedded watermark has a high robustness against signal modifications. In these applications, the embedded data is expected to have a high data rate and to be detected and decoded using a blind detection algorithm. While the robustness against intentional attacks is usually not required, signal processing modifications, like noise addition, should not affect the covert communications [2]. To qualify as steganography applications, the algorithms have to attain statistical invisibility as well. The algorithms presented in papers I-X were not designed to be statistically undetectable, thus the steganalysis of the algorithms is not in the scope of this thesis.

One interesting application of high capacity covert communications is public watermark embedded into the host multimedia that is used as the link to external databases that contain certain additional information about the multimedia file itself, e.g. copyright information and licensing conditions [2, 105, 106, 107, 108]. Another application with similar requirements is the transmission of meta data along with multimedia. Meta data embedded in, e.g. audio clip, may carry information about a composer, soloist, genre of music, etc. [105, 109].

Another possible application of high data rate information hiding schemes is audio streaming [105]. In many current audio-streaming applications, the audio bit stream is sent over the Internet using the TCP protocol. Supplementary data that contains information about audio content is, on the other hand, sent through the unreliable connectionless UDP protocol. As a result, the additional information is often lost in transmission, due to network congestion or router malfunction. Using audio data scheme, the need to use UDP for sending additional information can be circumvented by directly hiding additional

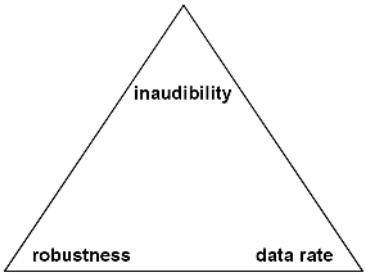

Fig. 3.1. **Magic triangle-three contradictory requirements of watermarking.**

information within the audio stream.

An additional application scenario is data hiding within analogue communication channels [105]. In order to hide data, analogue audio is sent thorough an analogue-to-digital (A/D) converter, and the output of the A/D converter is forwarded to the data hiding system. The output of the data hiding system is then fed through a digital-to-analogue (D/A) and modulated onto the analogue communications channel. The application is useful for users that want to receive extra data but do not have the requisite bandwidth for receiving the additional information. A high data rate covert communications system is able to transmit significant amounts of extra information for various applications.

3.1 High data rate information hiding using LSB coding

The algorithm that uses LSB coding is the natural choice of the watermarking algorithm that fulfils the requirements of high data rate and low robustness against signal modifications. It is one of the earliest and simplest steganography techniques and, as in cases of other known algorithms, it has first been developed for watermarking of images [49, 40, 54] and video stream [53, 110].

The watermark encoder uses a subset of all available host audio signal samples chosen by a secret key. The substitution operation on the LSBs is performed on this subset. The extraction process simply retrieves the watermark by reading the value of these bits. Therefore, the decoder needs all the samples of the watermarked audio that were used during the embedding process.

The main advantage of the method is a very high watermark channel capacity; the use of only one LSB of the host audio sample gives the capacity of 44.1 kbps if a mono audio signal, sampled at 44.1 kHz, is used. The obvious disadvantage is the method's extremely low robustness, due to the fact that the random changes of the LSBs destroy the coded watermark [53]. As no computationally demanding transformation of the host

signal needs to be done, this algorithm has a very small computational complexity. This permits the use of the LSB coding in real-time applications.

An increase in the embedding capacity is proportional to the number of the LSBs used for data hiding; two or more bits per sample could be used in order to enhance the bit rate of the hidden information. However, the increase of the number of the samples used during LSB coding introduces a low power additive white Gaussian noise. As already noted, the HAS is very sensitive to the AWGN and this limits the number of the LSBs that can be imperceptibly modified. In addition to a subjective quality degradation, the probability of the statistical detection of the embedded watermark increases as well [54, 55, 56, 110, 111].

There are two types of LSB insertion methods, fixed-size and variable-size embedding. The former embeds the same number of watermark bits in each sample of the host audio sequence. For the variable-size embedding method, the number of LSBs used for data hiding in each sample depends on the local characteristics of the host audio. It is still an open research issue how to adapt these local characteristics of the host audio in order to estimate the maximum embedding capacity.

3.1.1 Proposed high data rate LSB algorithm

The data hiding in the LSBs of audio samples in time domain is one of the simplest watermarking algorithms with very high data rates of hidden information. However, the adjusting of the LSBs of audio samples introduces noise that becomes audible as the number of the LSBs used for data hiding increases. An experimental test, performed in our laboratory, showed that for a large majority of music styles, three is the maximum number of the modified LSBs that leaves the watermarked audio perceptually transparent, if the host audio is represented with a 16 bits per sample resolution in time domain. Listening tests were carried out with a large collection of audio samples; furthermore individuals with a different background and musical experience took part. None of the tested audio sequences suffered audible perceptual distortion when 3 LSBs of its samples in time domain were used for data hiding. In addition, in certain music styles (loud rock or concert recording), the limit is even 4 or 5 LSBs per sample.

The embedding of additional information into consecutive LSBs injects AWGN to the levels that are above the JND level. Since the sensitivity of the HAS towards the additive random noise is high, a further increase of the watermark data rate using the standard LSB coding method is impossible.

We developed an advanced LSB coding method, which is able to shift the limit for transparent LSB data hiding from three to four LSBs, using a three-step approach (Paper V). Figure 3.2 illustrates the overall block-scheme of the proposed algorithm. In the first step, the algorithm embeds watermark bits to four LSBs of the host audio using a standard method, where the LSBs of the host audio's sample are simply replaced by four watermark bits. As noted above, in the majority of music styles, this causes a perceptual distortion of watermarked audio. Thus, some additional signal processing is needed in order to preserve the subjective quality of the watermarked audio.

Generally, if we embed k $(k < 16)$ bits in a sample, replacing the k LSBs of the sam-

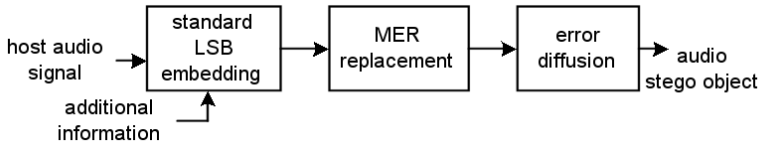

Fig. 3.2. Block-scheme of the proposed algorithm.

ple, the maximum embedding error introduced is $2^k - 1$. Considering 2^{16} levels of a 16-bit audio sequence, there are 2^{16-k} levels whose k LSBs are identical to the k embedded bits. In order to obtain the highest possible embedding transparency, the most similar value among these 2^{16-k} values should replace the original one. This is performed in the second step of the algorithm using a simple method to search for the level of audio closest to the original audio level as follows:
Let $a(n)$ be the original level of audio, $s(n)$ the level obtained by embedding k LSBs directly, and $s'(n)$ be the level of audio obtained by flipping the value of (k+1)th LSB of $s(n)$. The minimum-error level must be $s(n)$ or $s'(n)$. Let $e(n)$ be difference between $a(n)$ and $s(n)$ and $e'(n)$ be error between $a(n)$ and $s'(n)$. If $e(n) < e'(n)$, then $s(n)$ will be used to replace $a(n)$, otherwise $s'(n)$ is selected. This method is called a minimum-error replacement (MER) and has roots in high capacity image steganography algorithms [49, 112]. Using this method, we reduced the maximum embedding error from $2^k - 1$ to 2^{k-1}.

However, the loss of 6 dB of SNR that is introduced by increasing the number of the used LSBs by one cannot be compensated completely, because MER helps only in certain combinations of the incoming bits of information to be hidden. In order to decrease these perceptual artifacts, the third part of the algorithm is executed. This step has an error diffusion approach similar to improved grey-scale quantization (IGS), used for decreasing a false contouring in a quantized image, occurring due to an insufficient number of grey levels that would represent the smooth regions in the image [113, 114, 115]. In the digital image processing, the value of the embedding error is usually evenly spread to the bottom and right neighboring pixels, as shown in Figure 3.3. However, as audio signal is one-dimensional signal in time, an error caused by LSB modification can be diffused only in the "towards right", in other words, diffused to the samples that will be watermarked later. Let $e(n)$ denote the embedding error of the sample $a(n)$, then the next four consecutive samples of the host audio are modified according to: $a(n+1) = a(n+1) + e(n)/2$, $a(n+2) = a(n+2)+e(n)/4$, $a(n+3) = a(n+3)+e(n)/4$, $a(n+4) = a(n+4)+e(n)/8$.

The values that determine the distribution of embedding error to the consecutive samples have purposely been chosen to be a power of $1/2$. This means that all the weighting of the consecutive samples of the host audio is performed by a simple shift right operation where the number of shifts depends on the given weight. For example, the $a(n+3)$ sample is just shifted right for two positions and two zeros are written at the two most significant bits of the sample's binary representation. The weighting operation, performed in the

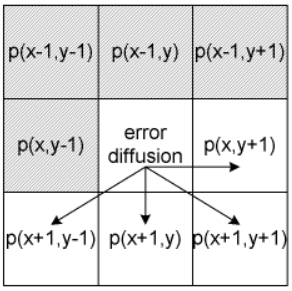

Fig. 3.3. **Error diffusion in improved grey-scale quantization used in image processing.**

given manner, facilitates a fast computation and keep the increase of the computational complexity of the overall algorithm minimal in comparison with the standard embedding method. All the modifications of the standard LSB algorithm are done at the embedding side while the extracting side carries the same computational burden. The increase in the number of computational operations will be executed by the main server in the multimedia distribution network. As it provides the multimedia content and performs data hiding, it has a far more computational power than receiving devices (laptops, PDAs, mobile phones, etc.). Therefore, the increase in computational complexity will not affect the end users.

The results of the subjective tests showed (Paper V) that the perceptual quality of watermarked audio, when embedding is done by the proposed algorithm, is higher in comparison with the standard LSB embedding. Test results indicated that a modified algorithm with four LSBs used for data hiding performs practically the same as the original LSB embedding algorithms with three LSBs used. This confirms that the algorithm in Paper V succeeds in increasing the bit rate of the hidden data for one third without affecting the perceptual transparency of the resulting audio signal.

Current storage requirements for digital mono audio signals are 705.6 kbps (sampling at 44.1 kHz and resolution 16 bits per sample). On the other hand, a reported perceptual entropy for wideband monophonic audio signals is in the range of 4-5 bits per sample [32, 116]. This implies that for an uncompressed audio signal, a significant amount of additional information can be inserted into the signal without causing a perceptual distortion. The theoretical bound is therefore from 485.1 to 529.2 kbps in data rate. The simple LSB coding method in time domain is able to inaudibly embed 3-4 bits per sample (132.3-176.4 kbps), which is far from a theoretically achievable rate, mostly due to a poor shaping of noise introduced by embedding and operation in time domain (Paper V). Therefore, a perceptual entropy measure of audio signals [116] and information theoretic assessment of the achievable data rates of a data hiding channel is necessary to develop a scheme that could obtain higher data rates.

3.2 Perceptual entropy of audio

It is a well-known fact, obtained during decades of audio compression research, that only a few bits per sample are needed to represent compact disk quality music. When performing a bit rate reduction of audio or speech signals that will be presented to the HAS, the objective is to introduce either imperceptible or inoffensive distortion during the compression process. This implies that for uncompressed music, noise can be injected into the host audio signal without being audible to the end user [32]. In audio steganography, this fact is used not for compression, but for embedding additional data. An estimate of the **perceptual entropy** of audio signals is created from the combinations of several noise masking measures. The results of tone-masking-noise and noise-masking-tone, as well as research on critical bands and spreading functions are combined in order to estimate the short term masking templates for audio signals [116]. The perceptual entropy of each short-term section of the audio signal is estimated as the number of bits required to encode the short-term spectrum of the signal to the resolution required to inject noise below the masking template level. When a bit rate reduction of an audio (or speech) signal is presented to the HAS, the objective is to introduce either imperceptible or inoffensive distortion during the compression process. The masking threshold for the audio

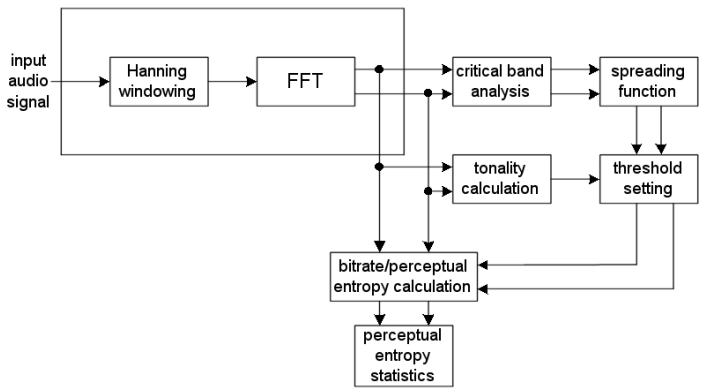

Fig. 3.4. **Perceptual entropy calculation algorithm.**

signal indirectly shows the amount of quantization that may be applied in the frequency domain, i.e., the quantization, according to the masking model, that may be done without corrupting the signal such that it can be distinguished from the original [116]. The part of the signal that can be modified without causing a subjective quality degradation is therefore perceptually redundant, and the part that must be preserved during the compression process represents real information that can be quantized and measured. In an ideal transform coder, the quantization step size and the number of levels in the quantizer

for each spectrum component could be set independently and without side information to communicate the level or bit allocations to the decoder. If the quantization step size in this ideal coder were set such that the total noise injected at each frequency corresponds to the threshold (the minimum number of quantization levels are used) then the number of bits required to encode the entire transform represents an estimate of the minimum number of bits necessary to transmit that block of audio. The total rate, divided by the number of samples coded, represents the per sample rate. The minimum per sample rate of this ideal transform coder needed to transparently encode an audio signal is called the perceptual entropy of the signal. This model is attractive, because it takes into account all of the artifacts and redundancies in the audio signal in the same manner as the HAS does (pitch, short term spectral model, etc.). There are three main parts of the perceptual entropy calculation algorithm [116], given in Figure 3.4:

1. Windowing of audio signal and transformation to Fourier domain
2. Calculation of the masking threshold
3. Calculation of the number of bits required to quantize spectrum of the signal.

The windowing of the signal is performed using a Hanning window and frequency transformation by FFT of length 2048. The first 1024 complex lines are kept (including the DC and lines counted as one line). The steps involved in calculating the masking threshold are critical band analysis, applying the spreading function to critical bands, calculating the spread masking threshold, accounting for absolute thresholds and, finally, relating the spread masking threshold to the critical band masking threshold.

3.2.1 Calculation of the perceptual entropy

As noted above, the perceptual entropy is calculated by measuring the actual number of quantizer levels to follow the signal in the frequency domain, given a step size in the quantizer that will result in noise energy equal to the audibility threshold [32]. Audibility threshold T_i is usually defined in the power domain and quantization energy is spread across k spectral lines in each critical band. It is also assumed that the quantization noise is spread uniformly across the entire critical band. The distribution of the quantization error is uniform in the amplitude domain; it gives noise variance equal to $\sigma^2/12$.

The step size S_i is calculated as follows. First, the energy is spread across the entire band, i.e. the energy at each spectral frequency is equal to T_i/k_i. Since the real and imaginary parts of the spectrum are quantized independently, the energy at each frequency must be divided in half, specifically the energy at each spectral component is $T_i/2k_i$. The noise energy, due to quantization is $\sigma^2/12$, therefore $\sigma^2/12 = T_i/2k_i$ and since $\sigma = S_i$ we obtain $S_i = \sqrt{6T_i/k_i}$, where S_i is the quantizer step size. This is done in each of the n critical bands:

$$N_{Re}(\omega) = \text{abs}\left(\text{nint}\left(\frac{Re(\omega)}{S_i}\right)\right), N_{Im}(\omega) = \text{abs}\left(\text{nint}\left(\frac{Im(\omega)}{S_i}\right)\right) \tag{3.1}$$

for each σ within the critical band i. The function abs$(\cdot)$ represents the scalar absolute value function and nint$(\cdot)$ a function that returns the nearest integer to its argument. $N_{Re,Im}(\omega)$ represents the integer quantized value of the each spectral line. Then, for each

ω, and individually for real and imaginary parts, $N_{Re,Im}(\omega)$ is altered as follows:
if $N_{Re,Im}(\omega) = 0$, then $N'_{Re,Im}(\omega) = 0$
if $N_{Re,Im}(\omega) \neq 0$, then $N'_{Re,Im}(\omega) = \log_2 (2N_{Re,Im}(\omega) + 1)$.
This operation assigns a bit rate of zero bits to any signal with an amplitude that does not need to be quantized, and assigns a bit rate of $\log_2$(number of levels) to those that must be quantized. If, for example, the integer number is 1, three levels (-1, 0, +1) are required to quantize the particular line. As the signs of different spectral lines are random, the sign information must be included. When no levels are necessary, the transmission of the sign bit is unnecessary as well, and a 0 is assigned to that line. The total bit rate is then calculated as:

$$\text{Total Rate} = \sum_{\omega=0}^{\pi} (N'_{Re}(\omega) + N'_{Im}(\omega)) \tag{3.2}$$

and the rate per sample (perceptual entropy) of the audio sequence is given by

$$\text{Perceptual Entropy} = \frac{\text{Total Rate}}{2048}. \tag{3.3}$$

The term *perceptual entropy*, used throughout this section, therefore indicates the 2048 sample perceptual entropy, regardless of the sampling rate or bandwidth of the signal. The block-to-block changes in perceptual entropy values increase as the effective window length decreases, but the mean and extreme values do not change significantly [116].

Reported perceptual entropy for wideband monophonic audio signals is in the range of 4-5 bps, taking into account all the spectral complexity, spectrum range and dynamic range requirements. This implies that for an uncompressed audio signal, a significant amount of additional information can be inserted into signal without causing a perceptual distortion. There is obviously a considerable gap between the currently available data rates for high capacity covert communications and theoretically obtainable data rates [52, 105, 107].

As noted above, a simple LSB coding method in time domain is able to inaudibly embed 3-4 bits per sample (132.3-176.4 kbps) of additional data, which is far from a theoretically achievable rate, due to the generation of AWGN caused by LSB embedding in time domain. Therefore, an information theoretic analysis of the capacity of information hiding channel is necessary in order to design a scheme that can offer higher data rates.

3.3 Capacity of the data-hiding channel

First we consider a simple data-hiding channel shown in Figure 3.5 [117, 118]. Here, $\mathbf{X} \sim \left(f_X(x), \sigma_x^2\right)$ is the message to be embedded, $\mathbf{Z} \sim \left(f_Z(x), \sigma_z^2\right)$ is the additive noise channel and $\mathbf{Y} \sim \left(f_Y(x), \sigma_y^2\right)$ is the received signal at the output of the channel. We also assume $\mathbf{X}$ and $\mathbf{Z}$ are independent, implying that $\sigma_y^2 = \sigma_z^2 + \sigma_x^2$. The channel capacity is given by:

$$C = \max_{f_X(x)} I(\mathbf{X}, \mathbf{Y}) = \max_{f_X(x)} h(\mathbf{Y}) - h(\mathbf{Y} \mid \mathbf{X}) = \max_{f_X(x)} h(\mathbf{Y}) - h(\mathbf{Z})[bits] \tag{3.4}$$

$I(\mathbf{X}, \mathbf{Y})$ is the mutual information between $\mathbf{X}$ and $\mathbf{Y}$. For a given statistics $f_Z(z)$ and σ_z^2, the entropy of $\mathbf{Y}$ should be maximized, $h(\mathbf{Y}) = -\int f_Y(y) \log_2(f_Y(y))dy$ [bits], us-

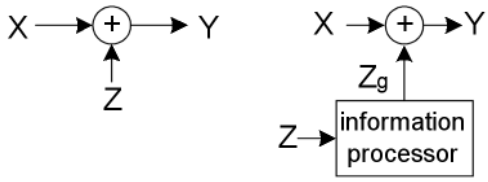

Fig. 3.5. (a) Simple data-hiding channel model, (b) Data-hiding channel model after Z is changed to a Gaussian distributed variable.

ing a suitable distribution $f_X(x)$ of the message **X**. For a given σ_y^2 the maximum value of $h(\mathbf{Y}) = \frac{1}{2}\log_2(2\pi e\sigma_y^2)$ bits is achieved when **Y** has a normal distribution. For instance, the maximum value of $h(\mathbf{Y})$ is achievable if both $f_Z(z)$ and $f_X(x)$ are normally distributed. However, for an arbitrary distribution $f_Z(z)$ and a fixed σ_x^2, the maximum achievable value of $h(\mathbf{Y})$ is not immediately obvious. This is because **Z** is usually altered in such a manner that the amount of information in **Z** is not altered, but the statistics of **Z** is changed to Gaussian distributed $\mathbf{Z}_g$. Therefore, for the purpose of calculating the channel capacity, we can replace $f_Z(z)$ by $N(0, \sigma_{zg}^2)$ and $h(\mathbf{Z}) = h(\mathbf{Z}_g) = \frac{1}{2}\log_2(2\pi e\sigma_{zg}^2)$ and we get:

$$C = \max_{f_X(x)} h(\mathbf{Y}) - h(\mathbf{Z}_g)[bits] = \frac{1}{2}\log_2\left(1 + \frac{\sigma_x^2}{\sigma_{zg}^2}\right)[bits] \tag{3.5}$$

The general data-hiding channel is usually decomposed into multiple channels, as hiding process is performed in a transform domain [117]. The decomposition is performed by the forward and inverse transform, as depicted in Figure 3.6. Signal decomposition into L bands results in L parallel channels with two noise sources in each channel. Let $\sigma_{ij}^2, j = 1, ..., L$ be the variances of the coefficients of each band of the decomposition. Let the corresponding Gaussian variances be σ_{igj}^2. If σ_{pj}^2 is the variance of the processing noise in the jth channel, the total capacity of the L parallel channels is given by:

$$C_h = \frac{N^2}{2L}\sum_{j=1}^{L}\log_2\left(1 + \frac{T_j^2}{\sigma_{igj}^2 + \sigma_{pj}^2}\right)[bits] \tag{3.6}$$

for a sequence of N samples. In the equation 3.6, T_j is the masking threshold of band j, in other words, the maximum power of the embedded message permitted in band j. In the case of no-processing noise (or if the processing noise is negligible), and we assume that all the channel have the same probability distribution function (such that $K\sigma_{ij} = K\sigma_{igj}$), the channel capacity is given by:

$$C_h = \frac{N^2}{2L}\sum_{j=1}^{L}\log_2\left(1 + \frac{K}{\sigma_{ij}^2}\right) \approx \frac{N}{2L}\log_2\left(1 + \sum_{j=1}^{L}\frac{K}{\sigma_{ij}^2}\right)[bits] \tag{3.7}$$

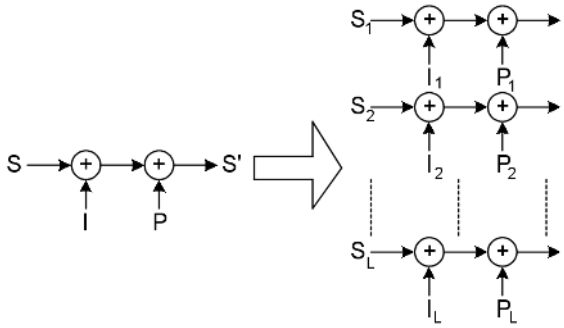

Fig. 3.6. **Decomposition of the data-hiding channel into multiple channels.**

It is clear that the **minimum channel capacity** is obtained when $\sigma_{ij} = \sigma, \forall j$ or when **no decomposition** is employed [118]. A transform with a good energy compaction or high gain of transform coding (GTC) [118] would result in more imbalance of the coefficient variances, resulting in an increased channel capacity. Therefore, a wavelet decomposition or discrete cosine transform (DCT) are good decompositions for low processing noise scenarios. The term *processing noise* here refers to equivalent additive noise which accounts for the reduction in correlation between the transform coefficients of the original signal and the transform coefficients of the audio signal obtained after MPEG compression, noise addition, low pass filtering, etc. On the other hand, the reduction in capacity with an increase of processing noise tends to be lower for transforms which are not used in compression methods, like DFT. While severe MPEG compression is certain to remove almost all high frequency components of DCT coefficients, it will not affect the high frequency DFT at the same extent. A signal decomposition with a low GTC is generally more immune to processing noise than decomposition with a high GTC and should predominantly be used in applications demanding robust watermarks. Therefore, signal decompositions with a high GTC, like the wavelet transform or DCT, are more suitable for high data rate steganography applications, where processing noise variance is low, because no intentional attacks are expected.

3.4 Proposed high data rate algorithm in wavelet domain

Using results from the information theory basis given above, we designed a novel audio steganography method with a high data rate of embedded information (Paper IV). The application scenario was to embed a MPEG compressed video sequence (high data rate requirement) into the host audio signal (mono signal, sampled at 44100 Hz). One example

of the practical implementation of the algorithm was the hiding of the artist's video clip in the artist's audio track (CD format). If the watermarked music clips are, e.g. compressed to the mp3 format, the embedded video clip can not be extracted. Therefore, no attacks or unintentional signal manipulations were expected, because it is the interest of the end user to obtain both multimedia files at the high quality data rate. The implemented method is a case of a fragile watermarking, as any distortion of the host audio signal leads to a severe quality loss of the embedded video clip.

Due to a low processing noise, the optimal selections of the signal decomposition algorithm are the wavelet decomposition and DCT. The wavelet domain is more suitable for frequency analysis because of its multiresolutional properties that provide access both to the most significant parts and details of signal's spectrum. Therefore, we are able to make easily the trade-off between the amount of the embedded information and perceptual distortion caused by information hiding, by handling subbands with different levels of power and perceptual significance.

Data hiding in the LSBs of the wavelet coefficients is practicable due to the near perfect reconstruction properties of the filterbank. The Discrete Wavelet Transform (DWT) decomposes the signal into low-pass and high pass components subsampled by two, whereas the inverse transform performs the reconstruction. We decided to make use of the simplest quadrature mirror filter - Haar filter. The Haar basis is obtained with a multiresolution of piecewise constant functions [36]. The scaling function is equal to one. As the equivalent filter has two non-zero coefficients equal to $2^{-1/2}$ at $n = 0$ and $n = 1$ Haar wavelet is defined as:

$$\psi(t) = \begin{cases} -1 & \text{if} \quad 0 \leq t < 1/2; \\ 1 & \text{if} \quad 1/2 \leq t < 1; \\ 0 & \text{otherwise.} \end{cases} \tag{3.8}$$

The Haar wavelet has the shortest support among all orthogonal wavelets, and it is the only quadrature mirror filter that has a finite impulse response [36]. FIR filters can be designed to be linear phase filters, which is important from the point of view of the perceptual transparency, as the linear phase filters delay the input signal, but do not distort its phase. In addition, the Haar filter is computationally simple to implement, as on most DSP processors, the FIR calculation can be done by looping a single instruction. This property gives the opportunity for real time applications of the proposed algorithm. FIR filters have also desirable numeric properties. In practice, all DSP filters must be implemented using a finite precision arithmetic and a limited number of bits. As FIR filters have no feedback, they can usually be implemented using fewer bits, and the designer has fewer practical problems to solve related to non-ideal arithmetic, in comparison with IIR filters [36].

Signal decomposition into the low-pass and high pass part of the spectrum is performed in five successive steps. After subband decomposition of 512 samples of host audio, using the Haar filter and decomposition depth of five steps, the algorithm produces 512 wavelet coefficients. All 512 wavelet coefficients are then scaled using the maximum value inside the given subband and converted to binary arrays in the two's complement. A fixed number of the LSBs are thereupon replaced with bits of information that should be hidden inside the host audio. Coefficients are then converted and scaled back to the original order of magnitude and an inverse transformation is performed. The details of the decomposition of the signal and subsequent data embedding are given in Figure 3.7.

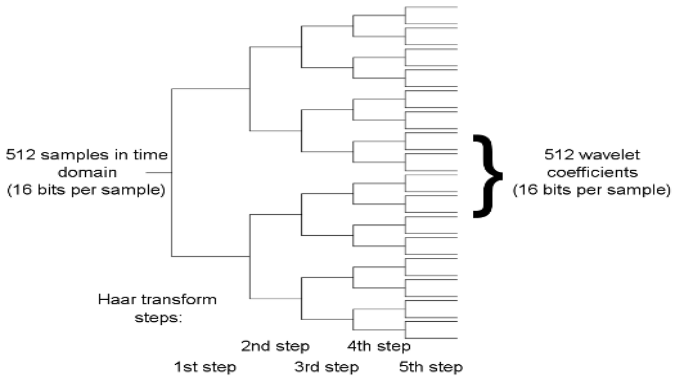

Fig. 3.7. **Signal decomposition prior to LSB embedding.**

The scheme was implemented using the integer wavelet transform (IWT) as well; in that case, there is no need for transforming coefficients (real values) into the integer format used for LSB embedding because IWT returns integers and would allow implementation on software with a less precise calculation than the Matlab© 16 bit floating point system.

The experimental results presented in the Paper IV are given for the case when wavelet coefficients of each of 32 subbands are modified in order to hide information. This is far from the optimal data hiding concept, as it has already been shown that the modification of the first four blocks of subband coefficients causes the largest degradation of perceptual quality of host audio [81, 119, 120, 121]. Nevertheless, we tried to make a balanced comparison between the proposed algorithm and the time domain LSB coding, for the case when we use the same embedding method and add noise to the host audio in all parts of audio spectrum. Some other simple solutions that would add to the performance of the proposed data hiding algorithm because the randomizing of input data and removal of the DC bias caused by LSB replacement are not used during the tests for the same reason.

During the subjective quality experiments (Paper IV), evaluation started with audio excerpts with three replaced LSBs for time domain and seven LSBs in wavelet domain because embedding to lower LSBs did not cause any noticeable perceptual distortion. The subjective experiments showed that the subband information hiding scheme has a large advantage over the classic LSB algorithm. The wavelet domain algorithm produces stego objects perceptually hardly discriminated from the original audio clip even when 8 LSBs of coefficients are modified, providing up to 5 bits per sample (220.5 kbps) higher data rate in comparison to time domain LSB algorithm.

The achieved bit rate of hidden information (Paper IV) is clearly above the bit rate obtained by other developed audio steganography schemes [52, 105]. In addition, the scheme can easily be modified to be more robust against processing noise (achievable bit rate would be decreased though) and it was used as a basis for the development of a robust

audio watermarking technique in wavelet domain [122].

3.5 Summary

Chapter 3 presented an insight in the first research subproblem of the thesis and the general background and requirements for high bit rate covert communications for audio. The subproblem was characterized by the following question: What is the highest watermark bit rate obtainable, under perceptual transparency constraint, and how to approach the limit?

Details and experimental results for the modified time domain LSB steganography algorithm were discussed. The results of subjective tests showed that the perceptual quality of watermarked audio, when embedding is done by the proposed algorithm, is higher in comparison with the standard LSB embedding. The tests confirmed that the described algorithm succeeds in increasing the bit rate of the hidden data for one third without affecting the perceptual transparency of resulting audio signal. However, the simple LSB coding method in time domain is able to inaudibly embed only 3-4 bits per sample, which is far from the theoretically achievable rate, mostly due to a poor shaping of noise introduced by embedding and operation in time domain. Therefore, a perceptual entropy and information theoretic assessment of the achievable data rates of a data hiding channel was necessary to develop a scheme that could obtain higher data rates.

A high bit rate algorithm in wavelet domain was developed based on these findings. The wavelet domain was chosen for data hiding due to its low processing noise and suitability for frequency analysis because of its multiresolutional properties that provide access both to the most significant parts and details of signal's spectrum. The experiments showed that the wavelet information hiding scheme has a large advantage over the time domain LSB algorithm. The wavelet domain algorithm produces stego objects perceptually hardly discriminated from the original audio clip even when 8 LSBs of coefficients are modified, providing up to 5 bits per sample higher data rate in comparison with time domain LSB algorithm.

4 Spread spectrum audio watermarking in time domain

One of the first audio watermarking algorithms that we developed (Paper I) is a time domain spread spectrum algorithm. It embeds a spread-spectrum-based watermark into an uncompressed, raw audio by slightly modifying the values of samples of the host audio in time domain. The main motivation was the development of an algorithm with a low computational complexity and with an embedding and extraction of watermarks in time domain. One of the most robust methods already developed for audio watermarking was a time domain algorithm [68]. Therefore, we tried not to use transforms, like DFT, or cepstrum transform that shift the host audio to transform domains and back to temporal domain consequently. It would definitely be hard to prove mathematically that watermarking in time domain gives smaller computational complexity in comparison with other, non-temporal algorithms because it is hard to compare complexity with each developed watermarking scheme. However, time domain algorithms have at least a lower implementation complexity and a smaller number of blocks in embedding and extraction algorithms.

4.1 Communications model of the watermarking systems

In order to describe the link between watermarking and standard data communications, the traditional model of a data communications system is often used to model watermarking systems. In Chapter 2, the basic components of a data communications system, related to the watermarking process, are highlighted. One of the most important parts of the communications models of the watermarking systems is the communications channel, because a number of classes of the communications channels have been used as a model for distortions imposed by watermarking attacks [123, 124, 125, 126]. The other important issue is the security of the embedded watermark bits, because the design of a watermark system has to take into account access that an adversary can have to that channel.

4.1.1 Components of the communications model

The main elements of the traditional data communications model are depicted in Figure 4.1. The main objective is to transmit a message **m** across a communications channel. The channel encoder usually encodes this message in order to prepare it for transmission over the channel. The channel encoder is a function that maps each possible message into a code word drawn from a set of signal that can be transmitted over the communications channel. The code word mapped by the channel encoder is denoted as **x**. It is common, as we deal with digital data and signals, that the encoder consists of a source coder and a modulator. The source coder removes the redundancy from the input message and maps a message into a sequence of symbols drawn from some alphabet. The duty of the modulator is to convert a sequence of symbols from the source coder into a signal suitable for transmission through a physical communications channel. It can use different modulation techniques such as amplitude, phase or frequency modulation.

The definite form of the channel encoder's output depends on the type of the transmission channel used in a particular model, but it is usually described as a sequence of real values, quantized to some arbitrary precision. In addition, we assume that the range of values of the channel encoder is limited in some way, usually by a power or amplitude constraint.

The signal **x** is subsequently sent over the communications channel, which is assumed to be noisy. The consequence of the presence of noise is that the received signal, conventionally denoted as **y**, is generally different from **x**. The extent of the change depends of the level of the noise present in the channel and is modeled here as additive noise. In other words, the transmission channel is modeled as adding a random noise **n** to the encoder's output **x**. At the receiver part of the system, the received signal, **y**, is forwarded, as the input signal, to the channel decoder which inverts the encoding process and attempts to correct for errors caused by the presence of noise. This is a function that maps transmitted signals into messages $\mathbf{m}_r$. The decoding process is typically a many-to-one function, so that correct decoding is possible even using noisy coded words [127, 128]. If the channel code is well matched to a given channel model, the probability that the decoded message contains an error is negligibly small.

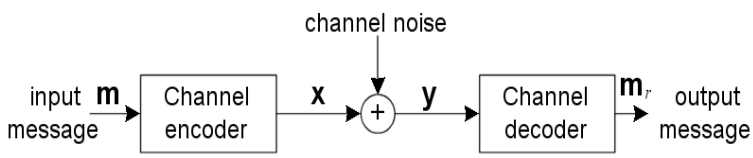

Fig. 4.1. **Standard model of a communications system.**

4.1.2 Models of communications channels

During the modeling of a communications system given in Figure 4.1, the parameters of the transmission channel are usually predetermined. That is, the function that is used for the modeling of the transmission channel cannot be modified during the transmission. The channel is generally characterized using a conditional probability distribution $P_{Y|X}(y)$, which gives the probability of obtaining **y** as the received signal if signal **x** was transmitted over the transmission channel.

Diverse communications channels can be classified in relation to the type of the noise function they apply to the signal and the way the distortion is introduced. The model from the Figure 4.1 is, as already mentioned above, an additive noise channel in which signals are distorted by the addition of noise signal **n**

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \tag{4.1}$$

The noise signal is usually modeled as independent of the signal **x**. The simplest and most important channel for analysis is a Gaussian channel where each element of the noise signal, $n(i)$, is drawn independently from a normal distribution with zero mean and a variance σ_n^2. The variance models the level of distortion of the signal introduced by channel noise and zero mean distribution means that channel noise does not have an impact on the DC component of the transmitted signal. Despite being simple, this model is the most frequently used one in the watermark literature and it was extensively used in our papers as well.

However, several non-additive communications channel models are also important. One of the frequently used models is the fading channel model [129] which cause the variation of the transmitted signal's power during the transmission. Generally, this variation can be modeled as a scaling of the signal

$$\mathbf{y} = v(t)\mathbf{x} \tag{4.2}$$

where $0 < v(t) < 1$ is an unknown parameter that vary slowly during the transmission or with each use of the channel. Such a channel might also include an additive noise component, rendering

$$\mathbf{y} = v(t)\mathbf{x} + \mathbf{n}. \tag{4.3}$$

There is only a small number of watermark papers that use a fading channel model for the description of the channel noise, one of the described models is given in Chapter 5 and Paper VIII.

4.1.3 Secure data communications

An important issue in watermarking is the security of the embedded watermark bits because the design of a watermark system has to take into account access that an adversary can have to the communications channel. In particular, we are interested in applications that demand security against passive and active adversaries. In the case of passive attacks, an adversary monitors the transmission channel and attempts to illegally read the message.

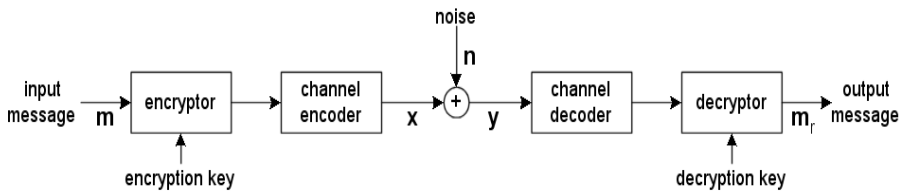

Fig. 4.2. A model of a communications channel with encryption.

In the active attack case, the adversary actively tries either to disable communication or transmit unauthorized messages.

There are two main methods of defence against attacks, as described in Chapter 2, first, cryptography and, second, spread spectrum communications. Prior to transmission, cryptography is used to encrypt a message using a secret key and after that the encrypted message is transmitted. On the receiver side, the encrypted message is received and then decrypted using the same or a related key to reveal the message. The block scheme is given in Figure 4.2. Cryptography introduces two advantages in a data communications system. The first is to prevent passive attacks in the form of an unauthorized reading of the message and the second is to prevent active attacks in the form of illicit writing. However, cryptography does not necessary prevent the adversary from knowing that a message is being transmitted. In addition, cryptography is helpless if an adversary intents to distort or remove a message before it is delivered to receiver.

Signal jamming (the deliberate effort by an adversary to inhibit communication between transmitter and receiver) was a great problem for military communications and has led to the development of the spread spectrum communication. In those systems, the modulation is performed according to a secret code that spreads the signal across a wider bandwidth than is regularly required. The code can be modeled as a form of the key used in the channel coder and decoder, as depicted in Figure 4.3. One of the examples of the spread spectrum communications is the frequency hopping method, one of the earliest and simplest spread spectrum techniques. In a frequency-hopping system, the transmitter broadcasts a message by first transmitting a part of the message bit stream on one frequency, the next fraction of the bit stream on the another frequency, and so on. A secret key that is known at the receiver as well as on the transmitter side controls the order of frequencies used for frequency hopping. Without a key, an adversary could monitor the transmission. The disruption of the transmission is also very difficult, because it could be done only by introducing noise at all possible frequencies, which would require too much power.

The cryptography and SS communications are complementary. The SS guarantees the delivery of signals, while the cryptography guarantees the secrecy of messages. Thus, it is common that these two technologies are combined in watermarking applications.

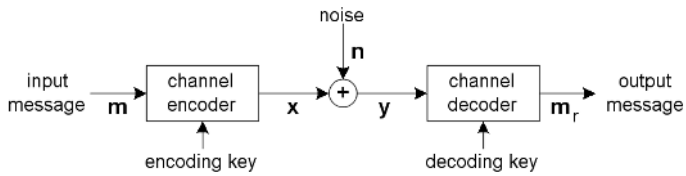

Fig. 4.3. **A model of a communications channel using spread spectrum key-based coding.**

4.1.4 Communication-based models of watermarking

The fundamental process in each watermarking system can be modeled as a form of communication where a message is transmitted from watermark embedder to the watermark receiver [2]. Therefore, it is natural to place watermarking into the framework of the traditional communications system. In Figures 4.4 and 4.5, two ways of mapping a watermarking system into communications framework are given. Figure 4.4 shows a watermarking system with an informed detection and Figure 4.5 a system that uses a blind detector.

In the watermarking-communications mapping, the process of watermarking is seen as a transmission channel through with the watermark message is being sent, with the host signal being a part of that channel. The embedding method consists of two basic steps,

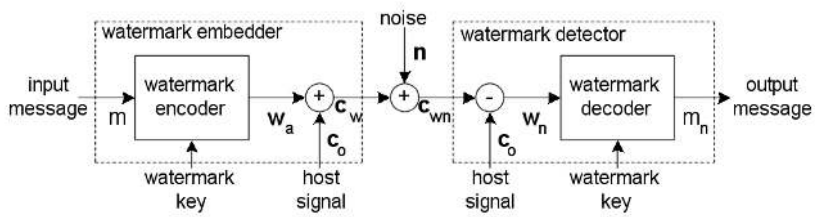

Fig. 4.4. **Watermarking system with informed detection-equivalent communications model.**

regardless of the detection method used (informed or blind detection). In the first step, the message to be transmitted is mapped into an added pattern, $\mathbf{w}_a$, of the same type and dimension of the host signal $\mathbf{c}_o$ (two dimensional patterns for images and videos and one dimensional patterns for audio). The mapping is usually performed using a secret watermark key. The calculation of the optimal added pattern $\mathbf{w}_a$ is typically performed in

several steps, and it starts with one or more reference patterns $\mathbf{w}_{r0}, \mathbf{w}_{r1}, \ldots$ which are predefined patterns, dependent on a watermark key. The reference patterns are subsequently combined to construct a pattern that encodes the message, which is referred to as a message pattern. The message pattern is the perceptually weighted in order obtain the added pattern $\mathbf{w}_a$. After that, $\mathbf{w}_a$ is added to the host signal $\mathbf{c}_o$, to construct the watermarked signal $\mathbf{c}_w$. If the watermark embedding process does not use information about the host signal, it is called the blind watermark embedding; otherwise the process is referred to as an informed watermark embedding. After the added pattern is embedded, the watermarked work is usually distorted during watermark attacks. We model the distortions of the watermarked signal as added noise, as in the data communications model. The types of attacks may include compression and decompression, broadcast over analogue channels, low pass filtering, dynamic compression, etc. However, the additive noise modeling is a simplified representation of the introduced distortions because all these types of distortions are non-stationary signal-adaptive processes.

If an informed watermark detector is used, the watermark detection is performed in two steps. In the first step, the unwatermarked host signal may be subtracted from the received signal $\mathbf{c}_{wn}$ in order to obtain a received noisy added watermark pattern $\mathbf{w}_n$. It is subsequently decoded by a watermark decoder, using the same watermark key used during the embedding process. Because the addition of the host signal in the embedder is exactly canceled by its subtraction in the detector, the only difference between $\mathbf{w}_a$ and $\mathbf{w}_n$ is caused by the added channel noise. Therefore, the addition of the host signal can be neglected, making watermark embedding, channel noise addition and watermark extraction equivalent to the data communications system given in Figure 4.3. In more advanced, informed detection systems, the entire unwatermarked host signal is not needed. Instead, some function of $\mathbf{c}_o$, usually a data reducing function, is used by the watermark detector to nullify "noise" effects represented by the addition the host signal in the embedder.

In a blind watermark detector, the unwatermarked host signal is unknown, and cannot be removed before a watermark extraction. Under these conditions, the analogy with Figure 4.3 can be made, where the added watermark is corrupted by the combination of impacts of the cover work and the noise signal. The received watermarked signal $\mathbf{c}_{wn}$, is now viewed as a corrupted version of the added pattern $\mathbf{w}_a$ and the entire watermarked

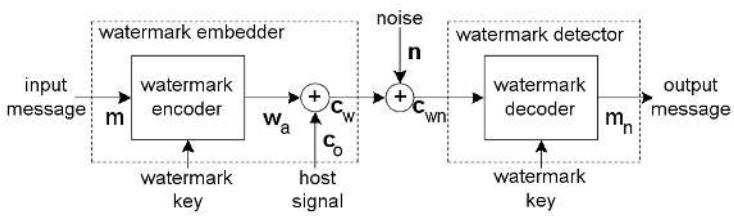

Fig. 4.5. A watermarking system with blind detection-equivalent communications model.

detector is viewed as the channel decoder.

In application that require robustness of the embedded watermark, e.g. a transaction tracking and copy control, the likelihood that the embedded message is identical to the extracted one, must be maximized, like in the traditional data communications systems. However, in the authentication watermarking systems, the goal is not to communicate a message, but to discover whether and how a host signal has been modified since watermark was embedded. Therefore, models from Figures 4.4 and 4.5 are not typically used to describe authentication systems.

4.2 Communications model of spread spectrum watermarking

A general model for spread spectrum-based watermarking is shown in Figure 4.6. Vector **x** is considered to be the original host signal already in an appropriate transform domain. The vector **y** is the received vector, in the transform domain, after channel distortions. A secret key K is used by a pseudo random number generator (PRN) to produce a "chip sequence" with zero mean and whose elements are equal to $+\sigma_u$ or $-\sigma_u$. The sequence **u** is then added to or subtracted from the signal **x** according to the variable b, where b assumes the values of +1 or -1 according to the bit (or bits) to be transmitted by the watermarking process (in multiplicative algorithms multiplication operation is performed instead addition [130]). The signal **s** is the watermarked audio signal. A simple analysis of SS-based watermarking, given in Chapter 2, leads to the probability of error equation for SS-based watermarking systems:

$$p = Pr\left\{\hat{b} < 0 | b = 1\right\} = \frac{1}{2}\text{erfc}\left(\sqrt{\frac{\sigma_u^2 N}{2(\sigma_x^2 + \sigma_n^2)}}\right) \tag{4.4}$$

where erfc(·) is complementary error function and the host audio **x** and the attack noise **n** are modeled as uncorrelated white Gaussian random processes: $x_i \sim N(0, \sigma_x^2)$ and $n_i \sim N(0, \sigma_n^2)$. It is clear that four parameters have an impact on the robustness of the

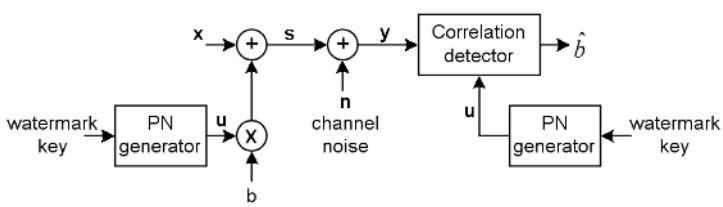

Fig. 4.6. A general model for spread spectrum-based watermarking system.

watermark detection process, power of the pseudo-noise sequence, length of vectors used for cross-correlation calculation, power of the host signal and power of the channel noise. The detection reliability increases with an increase in length of vectors N and the power of the pseudo noise sequence.

However, there are design limits in enlarging the power of chip sequence and length of correlation calculation. The increase in power of the chip sequence is limited by the requirement of perceptual transparency posed by the HAS. As already elaborated in Chapter 2, the HAS is very sensitive to the additive random noise in audio sequences, limiting the power of the added spreading sequence to a low level noise. On the other hand, an increase in the length of cross-correlation calculation does not have the impact on the perceptual transparency of the watermark system, but limits the capacity of the scheme. As N increases, more transform coefficients or samples in time domain are needed for embedding of one watermark bit and the bit rate of the embedded watermark is proportionally decreased [131, 132]. The channel noise parameter is set by an adversary that tends to disrupt watermark transmission and prevent its detection from the watermarked audio. The maximum value of the channel noise is limited by the requirement that the attacked watermarked audio remains perceptually acceptable to a human listener.

The modification of each coefficient can be small enough to be imperceptible, because correlator detector output still has a high signal to noise ratio to obtain low error detection, because it despreads the energy present in a large number of coefficients. *Direct sequence* spread spectrum systems spread the bandwidth of the information by a large factor called a processing gain G_p. The processing gain, expressed in dB, is determined by the length of vectors N

$$G_p = 10 \log N. \tag{4.5}$$

In order to obtain a satisfactory reconstruction of the embedded watermark in the decoder the spread-spectrum system has to provide sufficient processing gain. The spread spectrum method has proven to be, besides QIM [38], one of the most efficient ways to embed the watermark in a robust manner. The advantages of spread spectrum and quantization index modulation methods include:

1. Watermark detection does not require the original host signal
2. It is hard to extract the watermark using statistical analysis under certain conditions [128, 133].

However, as all block-based algorithms, spread spectrum method does not obtain a correct watermark detection, if the extracted watermark and the original pseudo noise sequence are not correctly aligned. The correlation calculation discussed above is reliable only if the detection chips are aligned with those used during embedding. Therefore, a malicious attacker can attempt to desynchronize the correlation by time- or frequency-scale modifications. There is a methodology for adding redundancy to the watermark chip pattern, called a redundant chip coding, so that the correlation metric is still reliable in the presence of scale modifications [33].

The basic idea behind redundant chip coding is shown in Figure 4.7. Figure 4.7(a) shows a perfect synchronization between a nine-chip watermark and a corresponding extracted watermark. The normalized correlation in that case totals $Q = 1$. However, if the watermark is shifted for one sample as in Figure 4.7(b), the normalized correlation equals

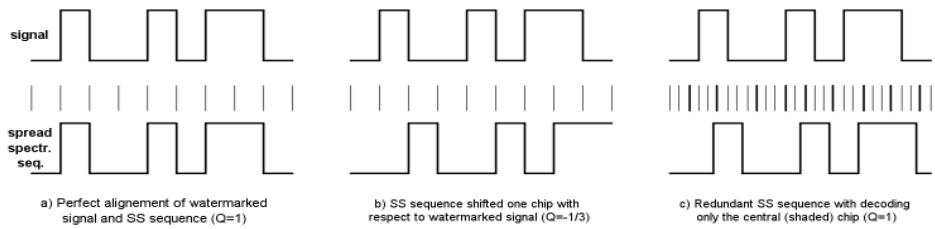

Fig. 4.7. **A reliable watermark extraction in the presence of scale modification attacks (Shaded time instances depict the time of cross correlation calculation for redundant chip coding).**

$Q = -1/3$. Thus, the detection process returns a negative decision, even though the signals are related. To prevent this type of an attack, each chip of the SS sequence is repeated in R consecutive samples, using redundant embedding. In this case, the trade-off between number of redundant repetitions, which decrease linearly the data rate of the embedded watermark, and robustness against desynchronization must be made. During the detection process, only the central sample of each R-tuple is used for computing the correlation. In our example in Figure 4.7(c), we use $R = 3$ which is sufficient to result in $Q = 1$. By using such an encoding and decoding scheme, it is straightforward to prove that the correlation is guaranteed to be correct even if a linear shift of $\lfloor R/2 \rfloor$ samples across the watermarking domain is induced. The issue of synchronization in spread spectrum watermarking schemes is still an open research issue, as resynchronization algorithms can offer only protection against a certain range of desynchronization attacks.

4.3 Spread spectrum watermarking algorithm in time domain

The basic audio watermarking algorithm that we developed is a time domain spread spectrum algorithm. It embeds a SS-based watermark into uncompressed, raw audio by slightly modifying the values of samples of the host audio in time domain. The procedure uses the virtues of the spread-spectrum communications given above, as well as temporal masking property of the HAS and the basic information about the spectrum of the host audio (Paper I). Figure 4.8 gives a general overview of the proposed watermark embedding algorithm. A simple trade-off between the watermark data rate and the robustness of the embedded watermark is possible, because the m-sequence length is decreased, the algorithm is able to embed a higher data rate watermark, but with less robustness against common watermark attacks, because low pass filtering or MPEG compression. For example, with the spreading sequence block length of 1023 samples, a watermark data rate of 43.10 bps is obtained.

The host audio sequence is initially analyzed in time domain, in order to determine the just noticeable distortion threshold, using the time domain masking property of the HAS. The goal is to place the watermark inside the host audio without causing a perceptual quality degradation in the process, while maximizing the amplitude values of the watermark sequence samples in order to increase algorithm's robustness in the presence of attacks. In the next step, a simple frequency analysis of the host audio is implemented as a common zero crossings counter in the basic block interval. The counting process derives information of the presence of the higher frequencies within the spectrum. If the presence of high frequency content is emphasized in a block, the power of the embedded watermark sequence can be greater as well, without affecting the overall subjective quality of the watermarked audio. The embedding algorithm obtains coefficient b(n) from the frequency analysis block, with higher values in the blocks in which the host audio has a significant high-frequency content. At the output of the watermark embedding process, the perceptually weighted spreading sequence is added to the host audio sequence resulting in:

$$y^*(n) = x(n) + a(n)b(n)w(n) \tag{4.6}$$

where a(n) and b(n) are coefficients obtained from temporal and frequency analysis blocks, respectively, x(n) is the host audio sequence and w(n) is the watermark sequence spread in time.

Figure 4.9 gives an overview of the watermark detection algorithm. The cornerstone of the detection process is, as in all spread spectrum systems, a cross-correlation calculation, in this case mean-removed cross-correlation between the watermarked audio signal and the equalized m-sequence (Paper I). Before the watermarked signal is segmented into blocks and cross-correlation with the m-sequence is calculated, the detection algorithm filters it with the equalization filter. The equalization filter is a high pass filter that filters out strong low pass components, increase correlation value and enhance detection results. The drawback is that it is a fixed coefficient filter, not adaptive to the local properties of the watermarked audio. The improvement of the detection robustness if adaptive filtering is used is presented in Section 4.5. The values from the correlation calculation block are forwarded to the detection/sampling block, which samples the output of the correlator

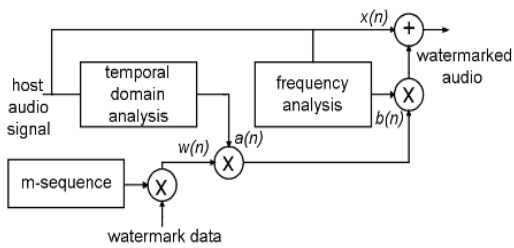

Fig. 4.8. **A proposed watermark embedding scheme.**

in order to obtain values for the threshold/decision block. The threshold/decision block provides the majority vote decision regarding the value of the embedded bit, depending on the sign of the correlation value.

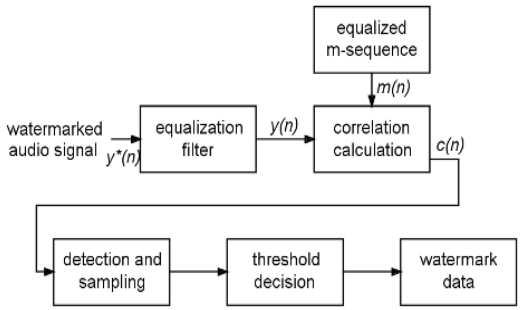

Fig. 4.9. **A watermark detection scheme.**

The correlation method, as already elaborated, demands alignment between the blocks of the equalized m-sequence and watermarked audio blocks in order to obtain reliable watermark detection. One of the malicious attacks on this scheme is the desynchronization of the correlation calculation procedure by time-scale modifications, such as the stretching of the audio sequence (without affecting the pitch) or the insertion/deletion of samples. In that case, the watermark detection scheme does not properly determine the value of the embedded watermark, resulting in a high increase of the bit error rate. A resynchronization algorithm that is able to provide a low bit error rate during the watermark decoding even in the presence of these attacks will be described in Section 4.4.

The algorithm obtained a high detection performance [123, 124, 125, 126] in the cases of band equalization, all-pass filtering, amplitude compression, echo addition and noise addition attacks (Paper I). After resampling and mp3 compression attacks, the bit error rate is higher than in the case of other attacks (Paper I), but the detection robustness was still equal to the other state-of-the-art algorithms. The reason for a poorer detection performance in the presence of a downsampling attack is that half of the spreading sequence power is lost after downsampling and strong low frequency components of the host audio remain unaffected by the attack. On the other hand, mp3 compression crops the high frequency spectrum of the watermarked audio and smoothes out audio waveform, destroying small modifications introduced by the watermark embedding algorithm.

The overall watermark detection robustness of the algorithm is comparable with other state-of-the art algorithms [72, 76], specifically in the presence of the most malicious attacks for SS watermarking algorithms (mp3 compression, resampling, low pass filtering). On the other hand, the algorithm uses computationally low demanding embedding and detection methods and a simple perceptual model for describing two masking properties

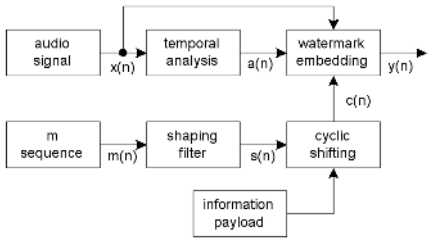

Fig. 4.10. **The improved watermark embedding algorithm.**

of the HAS. Thus, a successful compromise between the computational complexity and the detection performance of the algorithm is obtained.

4.4 Increasing detection robustness with perceptual weighting and redundant embedding

After the development of the basic audio watermarking algorithm for digital audio, described in Section 4.3, we improved the performance of the given method by utilizing more of the HAS properties and using a redundant embedding during watermark insertion (Paper II).

The basic idea is that the spectrum of the m-sequence is shaped in accordance to the HAS in order to make the watermark even more imperceptible. An integration function is added jointly with a synchronization scheme in the receiver to obtain a higher robustness against attacks. For handling time scaling attacks, a multiple chip embedding is used. With these enhancements, a considerably lower demand for computational power is attained, and better time-scaling resistance than with our earlier algorithm.

Figure 4.10 gives a general overview of the watermark embedding algorithm. Prior to further processing, the m-sequence is filtered in order to adjust it to masking thresholds of the HAS in the frequency domain (Paper II). The frequency characteristic of the filter is the approximation of the threshold in quiet curve of the HAS. Despite the simplicity of the shaping process of the m-sequence in frequency domain, the result is an inaudible watermark as the largest amounts of the shaped watermark's power are concentrated in the frequency sub-bands with a lower HAS sensitivity. A significant number of computational operations needed for the frequency analysis of audio, which have to be run in order to derive global masking thresholds in a predefined time window, are skipped, making this scheme appreciably faster. Although standard frequency analyses have more accurate data about the audio spectrum, the simulation tests done with selected audio clips showed a high level of similarity with the frequency masking thresholds derived from the masking

model defined in ISO-MPEG Audio Psychoacoustic Model.

A cyclic shifted version $c(n)$ of the shaped sequence $s(n)$ is used to achieve a multi-bit payload. Every possible shift is associated with a different information content and watermark bit rate is directly proportional to the length of the m-sequence (Paper II). Therefore, a simple trade-off between the embedded data size and robustness of the algorithm is obtained. The host audio sequence is also analyzed in the time domain, where a minimum or a maximum is determined in the block of audio signal that has the length of 7.6 ms. As the result of this analysis, the watermark samples are weighted by the coefficient $a(n)$ in order to be adjusted to psycho-acoustic perceptual thresholds in time domain.

Therefore, the watermark signal is embedded into a host audio using three time-aligned processes. In the first stage, the m-sequence has been filtered with the shaping filter, where a colored-noise sequence $s(n)$ is the output. Samples of the $s(n)$ sequence are then cyclically shifted, where the shift value is dependent of the input information payload. At the output of the watermark embedding scheme, the shifted version of $s(n)$, sequence $c(n)$ is being weighted and added to the original audio signal:

$$y(n) = x(n) + a(n)c(n) \tag{4.7}$$

where $x(n)$ denotes input audio signal and $a(n)$ are coefficients from the temporal analysis block. The addition of the $c(n)$ sequence in the embedding process is done redundantly in order to make the system resistant to time scaling attacks that tend to desynchronize the extraction process.

The diagram of the audio watermark detection scheme is shown in Figure 4.11. The detection process is again performed using the mean removed cross-correlation between the watermarked audio signal and the equalized m-sequence. Before the start of the integration process, which determines the peak and the embedded bit, the block power normalization part normalizes the energies of the output blocks from correlation calculations. The integration block sums the normalized output block from correlation detection and determines the peak and its position. The detection reliability depends strongly on the number of accumulated frames. In general, the trade-off is made between the time of integration and the amount of hidden data.

The extraction scheme uses redundancy in the watermark chip pattern, similar to the one described in [33]. The basic idea is to spread each chip of the shaped m-sequence onto R consecutive samples of watermarked audio. It has been proved that the correlation is correctly calculated even if a linear shift of $\lfloor R/2 \rfloor$ samples across the temporal or frequency domain is induced. However, there is a trade-off between the robustness of the algorithm and computational complexity, which is significantly increased by performing multiple correlation tests.

The test results showed that if attacks are performed by mp3 and AAC compression and time-scaling, the bit error rate is higher than in the case of other attacks, but the detection performance is still within the range of the state-of-the-art algorithms [72, 76]. The reason for poorer extraction capabilities after mp3 and AAC coding is that these compression techniques crop high frequency spectrum of the watermarked audio, where most of the watermark energy is situated. Time scaling is one of the most malicious attacks on the block-based watermarking algorithms, but the redundant spread sequence embedding solution reduced decoding BER in the presence of these attacks to an acceptable level. The penalty for an improved watermark decoding is a decreased bit rate of the embed-

ded watermark. However, the bit rate is still within an acceptable range for copyright applications.

4.5 Improved watermark detection using decorrelation of the watermarked audio

The watermarking methods presented in the two preceding sections use a matched filter technique based on the cross-correlation of the embedded PN sequence. The matched filter detection is optimal in the sense of SNR in the additive white Gaussian channel [2]. However, the host audio signal is generally far from the additive white Gaussian noise, which leads us to the optimal detection problem using a pre-processing of audio by the decorrelation of audio samples before detection. We proposed an audio decorrelation algorithm (Paper III) for a spread-spectrum watermarking that improves the robustness of the watermark detection and demonstrate a high resistance to attacks.

In a correlation detection scheme, used for watermark extraction process in spread-spectrum watermarking algorithms, it is often assumed that the host audio signal is white Gaussian process [134, 135, 136, 137]. However, real audio signals do not have white noise properties as adjacent audio samples are highly correlated. Therefore, the presumption for an optimal signal detection in the sense of signal to noise ratio is not satisfied, especially if extraction calculations is performed in short time windows of audio signal. Figure 4.12.a depicts a probability density function (pdf) of 5000 successive samples of a short clip of the watermarked audio signal. It is obvious that the pdf of watermarked audio is not smooth and has a large variance.

In order to decrease correlation between the samples of the audio signal, we use least squares Savitzky-Golay smoothing filters (with different polynomial order and window length), which are typically used to "smooth out" a noise signal whose frequency span is large [138]. Rather than having their properties defined in the Fourier domain, and then translated to the time domain, Savitzky-Golay filters derive directly from a particular formulation of the data-smoothing problem in the time domain. The Savitzky-Golay filters

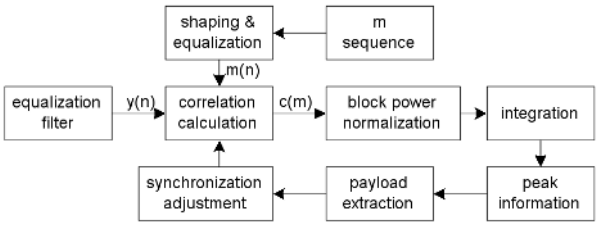

Fig. 4.11. **An improved watermark extraction algorithm.**

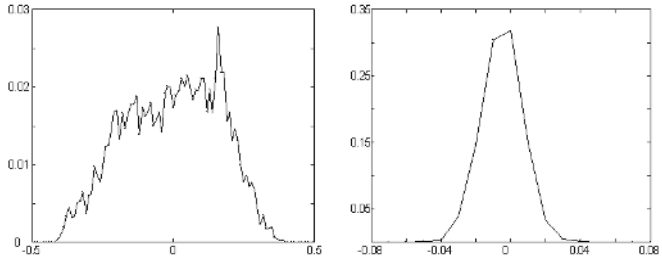

Fig. 4.12. **Probability density function of 5000 successive samples of a) watermarked audio signal b) watermarked signal after whitening process.**

are optimal in the sense that they minimize the least square errors in fitting a polynomial to frames of noisy data. Equivalently, the idea is to approximate the underlying function within a moving window by a polynomial, typically quadratic. Figure 4.12.b shows the pdf of the 5000 consecutive samples of the residual signal after applying Savitzky-Golay filters, with the fourth order polynomial and 21 samples long time windowing. It can clearly be seen that the pdf of the residual signal has a more Gaussian-like distribution and a significantly smaller variance compared to the case of the pdf of the watermarked audio signal. We verified a Gaussian-like distribution of the residual signal using the Bera-Jarque parametric hypothesis test of composite normality [139] and a single sample Lilliefors hypothesis test [140]. Both tests have rejected hypothesis that watermarked audio has Gaussian distribution, with a significance level of 5%. On the other hand, both tests also showed that we cannot reject the hypothesis that the residual signal has a Gaussian distribution, using the same significance level.

4.5.1 Optimal watermark detection

Pre-processed audio sequence **y** may have an embedded watermark

$$y(i) = s(i) + w(i), 0 \leq i \leq N - 1 \tag{4.8}$$

on the other hand, it may be an unwatermarked audio sequence

$$y(i) = s(i), 0 \leq i \leq N - 1. \tag{4.9}$$

The detection process verifies two hypotheses on the received content:

$\mathbf{H}_0$: watermarked audio content, so it is Gaussian white noise - residual signal of host audio after decorrelation process

$\mathbf{H}_1$: consists of decorrelated host audio and watermark

As decorrelation pre-processing was implemented, we can assume that the output of decorrelation filter **y** for a given **w** has the Gaussian distribution and the Likelihood Ratio Test can be performed.

In addition, the watermark part of the residual signal **w** is a sequence of samples w(i) with two equiprobable values, for example $w(i) \in \{-\epsilon, +\epsilon\}$ generated independently with respect to **s**. Parameter ϵ is set based on temporal analysis within one block of host audio. As the same PN generation and perceptual shaping of the PN sequence can be done on the receiver side, the correlation detector performs the simple correlation calculation between the pre-processed audio and whitened watermark sequence:

$$C = \mathbf{yw} = (\mathbf{s} + \mathbf{w}) \cdot \mathbf{w} = \mathbf{s} \cdot \mathbf{w} + \mathbf{w} \cdot \mathbf{w} = \mathbf{s} \cdot \mathbf{w} + N\epsilon^2 \tag{4.10}$$

where N is the cardinality of involved vectors, and the correlation between two vectors **a** and **b** is defined as $a \cdot b = \sum_{i=0}^{N-1} a(i)b(i)$. Since the host audio signal part of the residual audio clip **s** can be approximated as a Gaussian random vector $\mathbf{s} \sim N(\mu_x, \sigma_x), \sigma_x \gg \epsilon$, the normalized value of correlation can be written as:

$$Q = \frac{C}{N\epsilon^2} = \rho + \frac{1}{\epsilon} N\left(0, \frac{\sigma_x}{\sqrt{N}}\right) \tag{4.11}$$

where $\rho = 1$ if watermark is present and $\rho = 0$ if there is no watermark. The optimal detection rule is to declare that watermark is embedded in the host audio if the value of Q exceeds a given threshold value T. The selection of the threshold T controls the trade-off between a false alarm probability and the probability of detection. Using derivations from the Central Limit Theorem, probability that $Q > T$ is equal to:

$$\lim_{N \to \infty} P_r(Q > T) = \frac{1}{2}\text{erfc}\left(\frac{T\sqrt{N}}{\sigma_x\sqrt{2}}\right) \tag{4.12}$$

It is clear that the decorrelation of audio sequence leads to a decrease in variance value of signal σ_x (Figure 4.12), which again, according to the equations given above should lead towards a better detection performance and smaller false alarm probability [141, 142]. The dominant factor of the detection algorithm is determined by the autocorrelation of the whitened watermark sequences[143, 144], while the "noise" associated with audio covert communications channel is additive white Gaussian [145, 146].

The experimental results (Paper III) showed a significantly improved detection performance of the described method, compared to the standard watermark detection, if a watermarked audio sequence is attacked with an mp3 compression and low pass filtering attacks. The reason is that the attacked audio sequences still keep their amplitude-pdf different from Gaussian pdf. Therefore, the correlation detection is not optimal in the sense of Signal to Noise Ratio, because the channel can not be modeled as an additive white Gaussian noise channel. The residual signal has in both cases properties considerably more similar to AWGN and detection is accordingly more precise and stable. In the case of the amplitude compression attack, no significant improvement (Paper III) in detection results is achieved using a decorrelation filter, because the attacked audio already has a Gaussian-like pdf of amplitudes after an amplitude compression attack. In

general, the decorrelation algorithm improved the performance and stability of the watermark detection, because similar test results were obtained in the presence of other standard watermarking attacks, such as resampling, equalization and noise addition.

4.6 Increased detection robustness using channel coding

An equivalent model for watermarking is the process of data communications in which the goal is to successfully transfer the watermark data using information hiding techniques. In order to disrupt the communication stream, an attacker attempts to intentionally modify the watermarked signal in such a way that the watermark is removed, but the marked signal remains perceptually undistorted. The communication theory can be applied in order to find a relationship between the capacity of the watermarked channel and the distortion caused by a malicious attack. This section focuses on the problem of the watermark channel capacity, particularly on increasing the capacity of the watermark channel in the presence of attacks (such as low pass filtering and mp3 compression) by using turbo codes. The watermarking algorithm presented in Section 4.3 has the lowest detection reliability in the presence of mp3 compression, low-pass filtering and time scaling. Since an effective method resistant toward time-scaling attacks was already developed (Section 4.4), we decided to focus more on the low pass and mp3 attacks. As shown in [147], at the fixed signal to noise ratio, channel coding is the optimal solution for the decrease of bit error rate.

The watermark embedding scheme is the same as in Section 4.4. The watermark extraction part of the algorithm starts with a pre-whitening of the watermarked signal, described in Section 4.5. The correlator calculates a mean removed correlation between the residual signal $y^*(n)$ and pre-whitened PN-sequence $m(n)$. Correlation values follow a Gaussian distribution with a mean value μ and standard deviation σ, which depend on the type of music. Corresponding BER, using a hard limit decision, is therefore

$$BER = \text{erfc}\left(\frac{\mu}{\sigma}\right) = \text{erfc}(\sqrt{SNR}) \tag{4.13}$$

Values for BER without any attacks introduced increase as the capacity of the watermark channel increases. After introducing mp3 and LP attacks, BER dramatically increases. These attacks cannot be modeled as AWGN, due to the unpredictability of SNR variations (including complete fade) in the particular watermark channel during the watermark data transmission.

A far more appropriate model in this case is the frequency-selective fading model [129], because the fading model describes more precisely the distortion that appears when certain attacks are performed. For instance, in the algorithm described in Section 4.4, the watermark power is spread throughout the whole frequency range of audio and LP filtering crops all the spectrum components outside the pass band. Similarly, mp3 compression quantizes spectral components non-uniformly at different frequencies and it filters out the highest frequencies in order to preserve a level of perceptual fidelity.

4.6.1 Channel coding with turbo codes

In order to compensate for losses caused by attacks, we employ turbo codes (Paper VII) because they have a large coding gain and good properties in the fading channels [148, 149, 150, 151, 152, 153, 154, 155]. Similar improvement in detection results would probably be obtained if other channel codes were used. Turbo codes were chosen because of the level of expertise and developed software implementation the second author of the Paper VI has in the channel coding field. However, to facilitate turbo codes to produce a coding gain, the system must satisfy a minimal SNR condition, resulting in a decreased data rate of the watermark channel. The capacity of the watermark channel is defined as the maximum mutual information:

$$C = \max_{p(x)} I(X;Z) = \max_{p(x)} [H(X) - H(X|Z)] \tag{4.14}$$

where the maximum is taken over all possible distribution $p(x)$, X is watermark data after spreading and adjusting to the HAS properties and Z is the output from the watermark channel. The fading model of the watermark channel is given by

$$Z = G \cdot X + N \tag{4.15}$$

where G represents a random variable that models the channel fading variation and N is an AWGN with the variance $\sigma^2 = N_0/2$. The envelope amplitude of the fading attenuation G is a Rayleigh random variable. It is obvious that the channel capacity depends on whether the values of the fading attenuation G are known [147]; in this case, we do not estimate them. The penalty of not estimating channel state information (CSI) is around 0.8 dB for turbo codes that were used during the experiments (code rate R). It can be seen that the watermark bit rate is a trade-off between code rate and BER; the coding rate and watermark bit rate are directly proportional, while if we demand a lower BER the watermark channel capacity decreases. Therefore, the decreasing of the code rate will decrease the watermark data rate, but will also facilitate turbo codes to produce a lower BER for a fixed SNR per symbol and therefore increase the watermark channel capacity. This theoretical background gave us a solid foundation that introducing of turbo codes will reduce BER for a given watermark capacity in comparison with a regular detection or equivalently increase available watermark bit rate for a given BER.

The watermark bits are encoded before they are embedded into the host audio and iteratively decoded (Paper VII) using the soft output values from the correlator during the watermark extraction process. The watermark bits are divided in frames of 400 bits and encoded using multiple parallel-concatenated convolutional code. Interleaving inside frame was random and five decoding iterations of soft output values were performed in the turbo decoder. Each recursive systematic code was an optimum (5,7) code, giving a punctured code rate of $R = 1/2$. The frame length and code rate were chosen as a compromise between low computational complexity requirements of the watermarking algorithm and the demand for long iterations during turbo decoding process.

Test results showed (Paper VII) that turbo coding maintains a reliable watermark bit rate for a fixed BER, even after severe MPEG compression and filtering attacks. The watermark bit rate at fixed BER=10^{-6} is in the range of a few tens of bps (enough for the digital copyright applications), which was not attainable by the standard, uncoded

watermarking system. As expected [147], the uncoded system still slightly outperforms the one with turbo codec at low SNR per symbol values. Therefore, the introduction of the described turbo codec is justified only when the SNR per symbol value is high enough (spreading factor is large) and iterative decoding of soft output values is able to make the coding gain. One practical implementation issue could be the harsh slope of the watermark bit rate vs. BER curve (Paper VII), as a small change in the demanded bit rate causes a large BER variation. It can be simply solved by posing an upper limit for the BER value that will guarantee a certain range for the watermark bit rate.

4.7 Summary

Chapter 4 focused on the spread spectrum algorithms for digital watermarking and treats the second subproblem of the thesis. The subproblem was defined by the following question: How can the detection performance of a watermarking system be improved using algorithms based on communications models for that system? A general model for the spread spectrum-based watermarking is described as well, in order to place in context the developed algorithms.

A spread spectrum audio watermarking algorithm in time domain is presented. The overall watermark detection robustness of the algorithm is comparable with other state-of-the art algorithms, specifically in the presence of mp3 compression, resampling and low pass filtering. On the other hand, the algorithm uses computationally low demanding embedding and detection methods and a simple perceptual model for describing two masking properties of the HAS. One of the malicious attacks on this scheme is the desynchronization of the correlation calculation by time-scale modifications, such as the stretching of the audio sequence or insertion/deletion of samples. In that case, the watermark detection scheme does not properly determine the value of the embedded watermark, resulting in a high increase of the bit error rate.

A resynchronization algorithm that is able to provide a correct watermark detection even in the presence of these attacks, while maintaining a perceptual transparency by a perceptual noise shaping is presented subsequently. The consequence of the improved watermark decoding is a decreased bit rate of the embedded watermark; however the bit rate is still within an acceptable range for most copyright applications.

The possibility of improving the robustness of watermark detection and increasing the resistance to attacks was studied. An audio decorrelation algorithm for a spread-spectrum watermarking that uses least squares Savitzky-Golay smoothing filters is proposed. The test results showed a significant improvement in the detection performance of the described method, compared to the standard watermark detection, especially if the watermarked audio sequence is attacked with mp3 compression or low pass filtering attacks.

In order to further improve detection robustness and decrease bit error rate, the channel coding was employed, because it has property to reduce BER for a given watermark bit rate in comparison with a regular detection or equivalently increase an available watermark bit rate for a given BER. The simulations showed that the channel coding maintains a reliable watermark bit rate for a fixed BER, even after severe attacks. However, the introduction of the described turbo channel coding is justified only when the SNR value

is positive and the iterative decoding of soft output values is able to make the coding gain. One of the implementation issues was the harsh slope of the watermark bit rate vs. BER curve and the sensitivity to the cut attack, because the whole block of bits is needed during decoding.

5 Increasing robustness of embedded watermarks using attack characterization

As mentioned in Chapter 2, the main requirement of many watermark applications is the ability of the watermark detector to detect watermarks even if the watermarked audio has been significantly distorted after embedding. The watermarks embedded in such manner that they endure the legitimate and everyday usage of watermarked content are referred to as robust watermarks [2].

Recently, the watermark literature defined different types of robust watermarks. While the robust watermarks are designed to survive usual signal processing modifications, secure watermarks are designed to resist any attempt by an attacker to prevent their intended purpose [156, 157, 158, 159, 160]. As in most applications, the watermark system cannot perform its function if the embedded watermark cannot be detected, robustness is a necessary property if a watermark is to be secure. Therefore, if a watermark can be removed by an application of normal process it cannot be labeled as secure. On the other hand, robustness is not a sufficient condition for security, because secure watermarks must also survive processes that are specially designed to remove them. Thus, the design of a secure watermark system must take into consideration the range of all possible attacks, while the design of a robust watermark system can limit its focus to the range of probable processing.

Generally, there are several methods for increasing watermark robustness in the presence of signal modifications. Some of these methods aim to make watermarks robust to all possible distortions that preserve the perceptual quality of the watermarked signal. Others include strategies for enduring specific types of distortions. Some of the most frequent methods [2] for increasing robustness are:

1. Redundant embedding - watermark is redundantly embedded in several coefficients
2. Spread spectrum - redundant embedding strategy in frequency domain, already used in the design of robust audio watermarking systems described in Chapter 4.
3. Embedding in perceptually significant coefficients - modification of these coefficients to remove the watermark causes significant perceptual distortions of the watermarked media
4. Embedding into coefficients of known robustness - the modification is simulated at

the embedding side and the coefficients most resistant to it are selected for embedding process
5. Inverting distortions at the detector - during the detection process, the detector attempts to invert any processing that has been applied since the watermark was embedded
6. Pre-inverting distortions in the embedder - when there is a small set of distortions that watermark must survive, watermark is pre-distorted in order to be correctly detected.

In implemented watermarking systems, strategies for handling various types of distortions are usually combined. For example, image watermarking systems commonly use redundant embedding to handle cropping and noise addition, but use inversion in the detector to handle geometric distortions.

5.1 Embedding in coefficients of known robustness - attack characterization

When the watermark embedding is done in perceptually significant coefficients, the aim is to design a watermark that would survive all the possible attack that preserve a considerable level of perceptual quality of the attacked audio. However, in many applications the main focus is a specific set of attacks that might occur between the watermark embedding and detection. In such cases, the optimal approach is to deal with the specific attacks directly.

The first step is to find a domain of signal that is likely to be robust against the attacks of interest. For example, if we are more concerned with having an audio watermark survive temporal shifting than we are having it survive linear filtering, we might choose to embed in the FFT domain, because time domain shifting does not influence signal's spectrum. After the suitable domain for embedding has been selected, the coefficients that best survive the expected distortions are identified. The distortions that can be defined analytically allow the analytical derivation of the coefficients, for other distortions, it has to be done empirically. The experiments are generally straightforward and involve comparing the content directly after embedding and directly before detection. By comparing corresponding coefficients, we can find out how the channel between the embedder and the detector affects each coefficient. Such experiments need to be performed over a large number of samples, and numerous trials are often needed in order to get a suitable model with a sufficient statistical reliability.

However, a particular coefficient might be differently distorted in different host signals, like in the presence of adaptive compression. Adaptive compression algorithms, like mp3 compression, examine the signal to be compressed and set the amount of quantization applied to each coefficient. As a consequence, a particular coefficient can be heavily quantized in one audio signal, while almost unchanged in the other audio signal. This suggests that a watermark should be embedded adaptively.

One technique for determining the set of coefficients for individual host signal is to measure the relative robustness of each coefficient just prior to embedding a watermark. This is usually done by applying several simulated distortions to the host audio and measuring their effect on the coefficients of that work in the chosen domain. The watermark is

then embedded into the coefficients determined to be the most robust ones, which might be a different set of coefficients for each host signal. The subset of coefficients used for watermark embedding is forwarded to the detector along with the watermarked audio, which may be distorted. It is obvious that in this scenario an informed is required in order to extract watermark bits.

5.2 Attack characterization for spread spectrum watermarking

The primary goal of the introduction of the attack characterization into our audio watermarking algorithms was the poorer detection performance of the developed algorithms in the presence of mp3 compression, low pass filtering and resampling. The developed schemes had lower detection in the presence of time scaling (correlation desynchronization) attacks as well (Chapter 4), but a few algorithms have already been published [33, 78] that coped well with these watermark detection threats. Therefore, the main scope was the development of an attack characterization section in the embedding algorithm that would significantly improve detection results in the presence of frequency cropping attacks such as mp3 compression and low pass filtering. In addition, the design of an informed detector is needed in order to use data forwarded from the embedding side.

In spread spectrum watermarking, the embedded signal is a modulated low variance pseudo-random Gaussian white noise sequence. It is detected by cross-correlating the known watermark sequence with either the extracted watermark or the watermarked signal itself (informed or blind detection). If the correlation value is above a given threshold, then the watermark is detected. As elaborated in Chapter 2 and Chapter 4, the properties of the spread spectrum signalling makes it attractive for application in watermarking since a low-per-chip-energy, and hence imperceptible, watermark, robust to a narrowband interference, can simply be embedded and extracted.

However, the spread spectrum approaches have a number of limitations. For example, if the energy of the watermark is reduced due to fading-like distortions on the watermark, any residual correlation between the host signal and watermark can result in an unreliable detection. In addition, they neither take into account temporal nonstationarity of the host audio signal and attack interference nor include adaptive techniques to estimate the statistical variations. Furthermore, the correlator receiver structures used for the watermark detection are not effective in the presence of fading. Although spread spectrum systems in general try to exploit spreading to average the fading, the techniques are not designed to maximize performance. Many common multimedia signal distortions, including cropping, filtering, and perceptual coding, are not accurately modeled as narrowband interference. It has been proved [161, 162] that such signal modifications are fading-like on the watermark if embedded in an appropriate domain. The application of communication diversity and channel estimation techniques, which are effective in fading environment, is needed to obtain the robustness of watermarking schemes.

One of the earliest methods of attack characterization consisted of diversity and channel estimation [129]. Diversity is employed through watermark repetition and channel estimation through a reference watermark. Although it is well known that the repetition can improve the reliability of robust data hiding schemes, it is traditionally used to

decrease the effect of fading. If properly designed, a repetition can often significantly improve performance and may be worth the apparent sacrifice in the watermark bit rate. If the repetition is viewed as the application of communication diversity principles, it can be shown that a proper selection of an appropriate watermark embedding domain with an attack characterization can notably improve reliability.

5.2.1 Novel principles important for attack characterization implementation

There are three general principles used for the design of watermarking algorithms with an attack characterization [129], listed as follows.

1. Modeling of interference as fading
The previous analytic work in the area of robust digital watermarking has assumed additive Gaussian watermark channels. The effect of distortions on the overall watermarked signal and embedded watermark is considered to be in the form of stationary additive Gaussian noise. Intuitively, however, it is clear that some degradations such as cropping or heavy linear filtering have the effect of completely destroying the watermark content in the associated components of the signal. For example, if the watermark is embedded in the spectral domain of an audio signal, resampling the audio to a quarter of its original sampling frequency will destroy the watermark signal components in the discarded region of the signal while leaving others unchanged. Similarly, if the watermark is placed in the discrete Fourier transform components of the signal, a harsh low pass filtering will remove the existence of the watermark from high-frequency coefficients. Therefore, some very simple distortions have a nonuniform effect on the embedded watermark. That is, some watermark components are more severely distorted than others.

Fading is a term used to describe the effect of a communication channel that attenuates the information-bearing signal amplitude in an unpredictable way. Traditional characteristics of a general fading processing include:
• Varying SNR, including an SNR representing a complete fade of the watermark signal
• Unpredictability of SNR variations in the watermark channel before watermark transmission
• Independence of watermark signal attenuation in signal coefficients distant in frequency, time or another signal domain

2. Implementation of diversity
A general way to improve reliability in an unknown, nonstationary environment susceptible to deep fades is to employ diversity. A communication channel can be broken into independent subchannels, where each subchannel has a certain capacity. Since, in a fading environment, some of these channels may have a capacity of zero in a particular time instant, diversity principles are employed. Specifically, the same information is transmitted through each subchannel with the hope that at least one repetition will successfully be transmitted. For watermarking, it is referred to as *coefficient diversity* because different coefficients within the host signal are modulated with the same information. The sacrifice

in employing diversity is the bandwidth expense since the same information is sent using multiple coefficients.

3. Watermark channel estimation
In channel estimation, a training or reference sequence is employed to adjust the receiver filter to maximize the detection reliability. Watermarking methods that do not attempt to depict the attacks fail to exploit the advantage of extraction after any signal modification and, hence, fundamentally operate in a nonoptimal manner.The evaluation and demonstration of the performance improvements if watermark characterization is done prior to extraction is given in [129]. The analysis shown in [129] tries to find answers to two basic questions that are arisen when incorporating coefficient diversity and channel estimation:
• How to combine the different extracted repetitions of the watermark to maximize the overall reliability of the system?
• How to define sub-channels within the host signal to inherently promote robustness [162, 163, 164]?
The diversity and channel estimation should be incorporated into a general watermarking framework, e.g., through the use of a watermark repetition and attack characterization, respectively. Many proposed watermarking algorithms are encompassed by this class of techniques or can be easily modified to fit this category.

5.3 Watermark channel modeling using Rayleigh fading channel model

The first step in the development of the mp3 attack characterization is the estimation of channel distortions caused by mp3 compression. The analysis of the mp3 compression attack on the watermark channel (Paper VIII) was performed using a previously developed audio watermarking scheme, given in Paper II. Watermark is spread over a large number of samples in time domain and perceptual distortion is kept below the just noticeable difference level by using the occurrence of temporal masking effect of the human auditory system.

A pseudo random number generator is used to produce a "chip sequence" **u** with a zero mean and whose elements are equal to σ_u or $-\sigma_u$. We assume that one bit of information is embedded in a vector **y** of N samples in time domain, obtaining the watermark bit rate of 1/N bits/sample. A watermark bit is represented by the variable b, whose value is either -1 or +1. The watermark embedding is described by:

$$\mathbf{y} = \mathbf{x} + b\mathbf{u} \tag{5.1}$$

We assume a simple statistical model for the unwatermarked audio signal **x** - uncorrelated white Gaussian random process with a zero mean and variance σ_x^2. If there are no attacks, the normalized sufficient statistic at the detector follows a Gaussian distribution with a mean value μ_r and standard deviation σ_r, which depend on the type of music. In the case

when b=1 and a hard limit decision is used, the bit error rate (BER) p is given by:

$$p = \frac{1}{2} erfc\left(\frac{\mu_r}{\sigma_r\sqrt{2}}\right) = \frac{1}{2} erfc\left(\sqrt{\frac{\sigma_u^2 N}{2\sigma_x^2}}\right) \tag{5.2}$$

where *erfc* stands for the complementary error function. The same error probability is obtained if we assume that b=-1. The bit error probability, in the absence of attacks, increases as the bit rate of the watermark channel increases (smaller spreading N factor is used). If a watermark removal attack is introduced, watermarked signal at the detection is:

$$\mathbf{y} = \mathbf{x} + b\mathbf{u} + \mathbf{n} \tag{5.3}$$

where **n** is the noise caused by introduced attack. The BER in this case is given as:

$$p = \frac{1}{2} erfc\left(\frac{m_r}{\sigma_r\sqrt{2}}\right) = \frac{1}{2} erfc\left(\sqrt{\frac{\sigma_u^2 N}{2(\sigma_x^2 + \sigma_n^2)}}\right). \tag{5.4}$$

In the previous work in the field of watermarking, the noise introduced by attacks was usually modeled as Additive White Gaussian Noise (AWGN). The frequency analysis of the watermarked signal showed unpredictability of noise variations (including complete fade) in the particular frequency sub bands in the presence of mp3 coding. For instance, in the tested algorithm, the watermark power is spread throughout the whole frequency range of audio (Paper VIII). Imposed mp3 compression quantizes spectral components non-uniformly at different frequencies and filters out the highest frequencies in order to preserve a level of perceptual fidelity. In addition, it has been proved [147] that correlator receiver schemes are not very effective in the presence of a fading-like interference. Therefore, a far more appropriate model for the watermark channel in the presence of mp3 coding should be the frequency-selective fading model because it describes more precisely the distortions that appear when mp3 attacks are introduced. We assumed the Rayleigh frequency-selective fading channel model and that receiver does not have channel state information (CSI). The Rayleigh fading channel model was adopted as it is one of the simplest fading models. Therefore, if this model describes the attack distortions better than the standard model, it can be expected that more complex models would perform even better. Theoretical bit error rate for the Rayleigh fading channel is given by [147]:

$$p = \frac{1}{2}\left(1 - \sqrt{\frac{\frac{N\sigma_u^2}{2\sigma_x^2}}{1 + \frac{N\sigma_u^2}{2\sigma_x^2}}}\right) \tag{5.5}$$

In order to practically check the hypothesis, we compared the expected theoretical figures for BER derived from equations 5.4 and 5.5 with the BER curves obtained from experiments (Paper VIII). The experimental values of BER were obtained using a large set of watermarked songs from different music styles. The watermarked sequences have then been attacked using mp3 compression.

Our experiments suggest (Paper VIII) that the noise introduced by mp3 compression can hardly be modeled as AWGN, as BER curves differ as much as one order of magnitude for some values of the spreading factor N. The BER curves obtained by the Rayleigh

fading channel model have steepness and values more close to the experimentally derived ones. The results confirmed that a far better watermark channel modeling is obtained by the proposed model than with the usual AWGN watermark channel model.

5.4 Audio watermarking algorithm with attack characterization

Using the theoretical background from Section 5.1 and Section 5.2, we developed a novel audio watermarking scheme that uses an attack characterization to obtain high robustness against standard watermark attacks (Paper VI). The watermark embedding and detection are based on the frequency hopping spread spectrum method in the spectral domain.

The watermark embedding scheme is given in the Figure 5.1. Samples of the host audio sequence are forwarded to the SYNC module (Figure 5.1). In the SYNC module, the host audio is divided into blocks used for data hiding and blocks used for the watermark extraction synchronization. The data hiding blocks have a fixed length L, while synchronization blocks have a length chosen randomly from the interval $[L_1, L_2]$. Thus, between each two consecutive data hiding blocks, there is one synchronization frame with variable length. In each synchronization frame, a perceptually shaped PN sequence is added to the host signal in time domain. The spreading gain of the embedded PN sequence is controlled through the limits of the synchronization block length L_1 and L_2. The data hiding

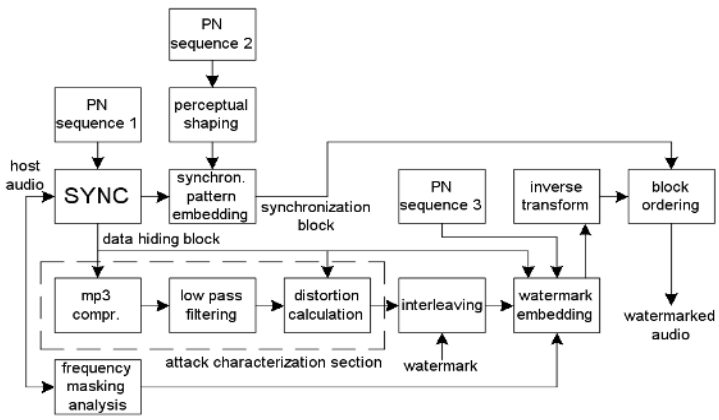

Fig. 5.1. **A watermark embedding scheme.**

block is forwarded to the attack characterization section of the embedding scheme (Paper VI). Each data-hiding block undergoes mp3 compression and LP filtering and distortion measure D, for the ratio of the original magnitude of an FFT coefficient and magnitude of

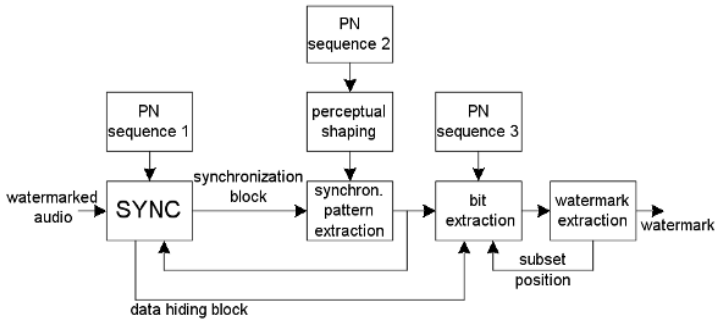

Fig. 5.2. **Watermark extraction method.**

the same FFT coefficient after modifications, is calculated during predefined time interval T. The algorithm selects a sub band corresponding to 100 consecutive FFT coefficients (of 1024 coefficients in total) with the least distorted magnitudes. At the embedding module, the binary coded identity of the position of the first coefficient is inserted along with watermark bits into single bit stream and embedded into data hiding blocks with a N-fold repetition during time interval T. The time interval T is chosen as a trade-off between two conflicting requirements. The first requirement is to get precise information about the distortion of FFT coefficients at a particular time instant, and the second one is decreasing the portion of the position identity bits in the unified data stream.

Data embedding is performed by a frequency hopping method. A secret key is used to select two FFT coefficients from the sub band least affected by modeled attacks. The mean value of the magnitudes of all the coefficients in the sub band is calculated and assigned to the two mapped coefficients' magnitudes. The magnitude of the coefficient at the lower frequency is then increased by K decibels (dB) and the value of the second coefficient is decreased by the same value, if bit 1 is to be embedded. The opposite arrangement is done if bit 0 is signalled. The value K is chosen to be equal to distance from the mean value of the magnitudes of the sub band to the frequency masking threshold, derived from the frequency masking property of the HAS. After the additional data bit has been embedded, the block is transformed back to the time domain and inserted between two synchronization frames.

At the start of the watermark extraction processes (Figure 5.2), samples of the watermarked audio are checked for synchronization. Mean removed cross-correlation, between synchronization block and the same prefiltered PN sequence as the one used during watermark embedding, is calculated. If a time shift is noticed, the following data hiding block is shifted for the same number of samples, after which the extraction process from the data hiding block begins. Using the same hopping key-based pattern as on the embedding side, the detector reads the magnitude (in dB) of the first FFT coefficient (value A); the same operation is repeated for the FFT coefficient on the higher frequency, obtaining

value B. The detection value V is calculated as the difference between values A and B. The sign of V determines the value of the extracted bit; a positive value is mapped to bit 1, otherwise bit 0 is extracted. After the time interval T, a new sub band is selected using the extracted information about the position of the first coefficient of the sub band.

The detection performance of the algorithm (Paper VI) was tested against the standard audio watermarking attacks. The algorithm showed a high performance in the presence of the amplitude compression, resampling and mp3 compression. Although the bit error rate (BER) was slightly higher with an echo addition and time scaling, it was still within the range obtained by other state-of-the-art algorithms. The detection results were compared with the results obtained using the same scheme without an attack characterization section (Paper VI). The results indicate that an attack characterization significantly improves the detection performance of the algorithm, decreasing the bit error rate 4 to 20 times in the case of LP filtering or mp3 compression attacks (Paper VI).

5.5 Improved attack characterization procedure

As noted in Section 5.4, all the contemporary SS audio watermarking algorithms have significantly decreased the detection reliability in the presence of low pass (LP) filtering and MPEG compression. These two attacks cannot be modeled as Additive White Gaussian Noise (AWGN) due to the unpredictability of SNR variation in the particular watermark channel during watermark data transmission. If the watermark power is spread throughout the whole frequency range of audio and LP filtering is introduced, watermark components outside pass band are significantly distorted. Similarly, MPEG compression quantizes spectral components non-uniformly at different frequencies; in addition, it filters out the highest frequencies in order to preserve a level of perceptual fidelity. Therefore, an improved technique must include a characterization of fading-like distortions of coefficients where the watermark is to be embedded and concentration of watermark energy in regions that are less distorted (Paper IX).

We developed a novel scheme that has a significantly higher detection robustness compared to the standard SS watermarking algorithm that uses direct sequence (DS) approach (Paper IX). Using the frequency hopping (FH) method [165], the good properties of the SS methods remain intact. In addition, there is no calculation of cross-correlation between the embedded SS sequence and host audio as in the standard SS algorithms, as the correlation calculation is replaced by a modified patchwork algorithm [122] at the extraction side. The watermark embedding scheme is similar to the one described in Paper VI, with two novel features. The scheme described in Paper VI used both MPEG compression and LP filtering attack characterization in order to find the subset of FFT coefficients least affected by these fading-like distortions. However, the experimental tests showed that the characterization section selects similar subsets of FFT coefficients (Paper IX) even if we leave out the LP filtering module, as the MPEG compression has an inherently embedded LP filter.

Therefore, for the reason of the decreased computational complexity of the embedding algorithm, only MPEG compression is simulated at the characterization section. The distortion measure D for the ratio of the original magnitude of an FFT coefficient C_i and

magnitude of the same FFT coefficient after the simulated attack C_i^*, is calculated during a predefined time interval T:

$$D = \sum_{i=1}^{N} a_i D_i, D_i = \frac{(C_i - C_i^*)^2}{C_i^2} \tag{5.6}$$

and $a_i = \frac{log(i+1)}{i}$ for $i = 1, \ldots, N$. Coefficients a_i are introduced because the experiments showed that the modification of the FFT magnitudes at the lower frequencies introduces more perceptual distortion, as they contain more signal energy. The a_i expression is derived from experimental data. Other models for weighting coefficients have been tested, with similar results; however, the experiments are done using the expression above. Subsequently, weights a_i improve the perceptual transparency of the algorithm, allowing less distortion in the frequency subbands of the higher sensitivity of the HAS.

The watermarking extraction scheme is identical to the one in Paper VI. If a time scaling attack is performed, the correlation peak is decreased for a random value, depending on the place where the samples of the watermarked audio were deleted or additional samples inserted. However, the parameters of the synchronization block enable a reliable detection of the correct position of the data hiding block, if the scaling factor is in range $[-3\%, +3\%]$. A further increase or decrease of the length in the watermarked audio significantly decreases the performance of the watermarking extraction scheme.

In order to make comparison with DS spread spectrum watermarking algorithms, we used one of the standard DS algorithms in FFT domain [76], with an embedding and extracting scheme given in Figure 5.3 and 5.4, respectively. The parameters of the DS algorithm were selected in such a way that the watermark bit rate is equal to the bit rate of our algorithm. The forwarding of the selected subset information to the watermark detector is done using the same method as in our algorithm.

The robustness of the algorithms was tested against the standard audio watermarking attacks listed in Paper VI. The results in the case of no attack characterization used at the watermark embedding scheme were obtained as a reference value as well. The experimental results proved a significant advantage in the detection robustness that the proposed algorithm has (Paper IX), compared to the DS spread spectrum algorithm, with a BER generally 4-10 times lower. In addition, it is clear that the introduction of the attack characterization module additionally improved the extraction reliability of both algorithms, decreasing the bit error rate, most discernibly in the presence of MPEG compression, low pass filtering and resampling attacks. Therefore, the algorithm obtained a high detection robustness, while decreasing the computational complexity and increasing the perceptual transparency of the watermarked signal.

5.6 Attack characterization section in an improved spread spectrum scheme

In [39] the authors describe the importance of decreasing the influence of the host signal on the watermark extraction process, analyzing a spread spectrum system with the fixed cross correlation value. The analysis of the watermark detection performance clearly

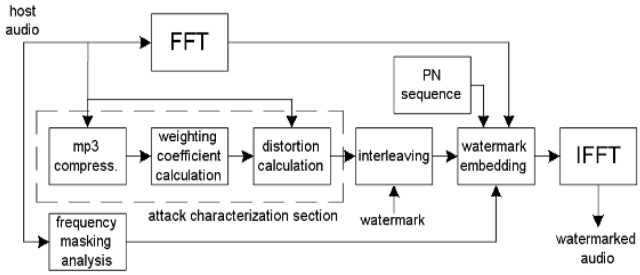

Fig. 5.3. **A direct sequence watermark embedding scheme.**

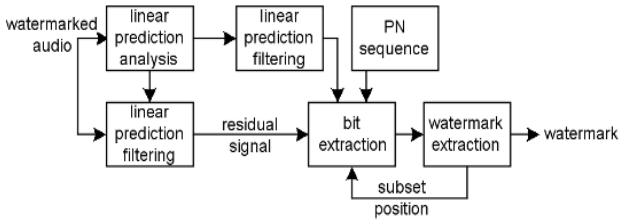

Fig. 5.4. **A direct sequence watermark extraction scheme.**

shows an improved detection robustness, in comparison with the case of an uninformed watermark embedding, where the host signal itself is considered as a source of interference in the watermark channel. However, in [39] there is no detailed description of the practical issues concerning the watermark embedding process, e.g. the control of the perceptual quality of the signal when a fixed cross-correlation is forced.

Using the framework from [39], in [90] authors derived three different watermarking approaches, corresponding to the cases of "maximized robustness", "maximized correlation coefficient" and "constant robustness". Still, the problem of minimizing the bit error rate, at a fixed average distortion level during the watermark embedding process, is not addressed. Recently, an improved spread spectrum (ISS) method has been proposed [91] that removes the host signal as a source of interference, gaining significantly on the robustness of watermark detection. The improvement obtained using ISS over standard SS method is in the range of gains if the quantization index modulation (QIM) is compared to the standard SS methods. The ISS method does not suffer from the same sensitivity to am-

plitude scaling as the QIM method, because ISS is insensitive to the amplitude scaling as the SS method. However, the ISS method cannot keep the distortion caused by watermark embedding at a constant level as in the SS method. Although it delivers the same average distortion as in the SS method, a forced cross-correlation minimization may cause a large local distortion of the host signal, which is an unacceptable property for most of audio watermarking applications. In addition, all the results presented in [91] are theoretically derived, without a subjective test and measuring the bit error rate in the presence of the attacks other than AWGN.

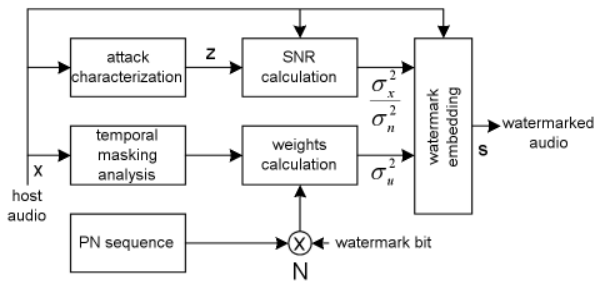

Fig. 5.5. **ISS watermark embedding algorithm**

We proposed a novel robust audio watermarking algorithm in time domain that uses the perceptually tuned ISS method and attack characterization at the embedding side (Paper X). The overall scheme of the watermark embedding algorithm is given in Figure 5.7. The samples of the host audio sequence are forwarded simultaneously to the masking analysis module and attack characterization module. The masking threshold in time domain is derived for every input block of host audio. The length of the frame and power level of watermark are chosen in line with the requirements of the HAS regarding inaudibility and to give the watermark highest possible amplitude before it is added to the host signal.

The attack characterization section has the purpose of the analysis of the signal for the watermark removal attacks including mp3 compression and LP filtering. In order to find the level of the introduced noise by these distortions, these spectrum modifications are simulated at the embedding side, where each data hiding block undergoes mp3 compression and LP filtering (Paper X). A distortion measure SNR is defined as:

$$\text{SNR} = 10 \cdot \log_{10} \frac{\sum_n x^2(n)}{\sum_n [x(n) - z(n)]^2} [dB] \tag{5.7}$$

is calculated for the blocks of host audio with a predefined length N and forwarded to the watermark embedding block. $x(n)$ stands for the original host audio samples and $z(n)$ are the samples of audio after the given modification.

The watermark bits are perceptually tuned using weight coefficients form the HAS time domain masking analysis and embedded into the host audio sequence using ISS modulation. The power of the watermark sequence in a block with length N, after spreading and perceptual tuning, is σ_u^2. We used the linear version of the ISS method, because it is the simplest to analyze, but still provides a significant part of the gains in relation to the traditional SS method. In this case, the host audio is watermarked according to:

$$\mathbf{s} = \mathbf{x} + (\alpha b - \lambda x)\mathbf{u} \tag{5.8}$$

where **x** stands for the original host signal vector, **s** stands for watermarked audio vector and **u** holds for the PN sequence after the perceptual adaptation process. A weighted PN sequence is added or subtracted from the signal **x** according to variable b, where b can be either +1 or -1, according to the watermark bit embedded into the host audio. Parameters α and λ control the distortion level and removal of the host signal influence on the detection statistic, respectively. Using the framework given in Section 2.3.5, it is possible to derive optimal values for λ and α:

$$\lambda_{opt} = \frac{1}{2}\left(\left(1 + \frac{\sigma_n^2}{\sigma_x^2} + \frac{N\sigma_u^2}{\sigma_x^2}\right) - \sqrt{\left(1 + \frac{\sigma_n^2}{\sigma_x^2} + \frac{N\sigma_u^2}{\sigma_x^2}\right)^2 - 4\frac{N\sigma_u^2}{\sigma_x^2}}\right) \tag{5.9}$$

$$\alpha = \sqrt{\frac{N\sigma_u^2 - \lambda^2\sigma_x^2}{N\sigma_u^2}}. \tag{5.10}$$

The watermark embedding scheme uses equation 5.9 for λ_{opt} for the adjustment of the desired properties and the overall performance of the watermarking system. The attack characterization module can include several sections that would simulate expected attacks that appear in the transmission channel. The test results are obtained using the attack characterization module that consisted of mp3 and low pass filtering characterization sections, because they caused the largest bit error rate on the original SS watermarking system (Paper II), as well as on other contemporary audio watermarking methods. The masking analysis module computes the highest allowed value σ_u^2 for under the constraints of time domain masking of the HAS. The estimate of the signal-to noise ratio in the watermark channel from the attack characterization block is forwarded to the embedding module.

Using the attack characterization, even by a simple parameter as SNR, we were able to implement a watermarking system that is able to make a trade-off between good statistical properties of ISS modulation and requirement for a robust watermark detection. As we aim to improve an algorithm using a blind watermark detection (without access to the original host audio sequence), it is a convenient way to estimate the channel noise **n** without the knowledge of the statistical model of the noise. For a desired watermark bit rate, determined by variable N, λ_{opt} is calculated and the variable a derived from the equation 5.9. Therefore, using the attack characterization block we can derive upper bounds for a system's performance under a particular watermark removal attack and determine the upper bound for the capacity of the watermark channel for a given bit error rate. On the other hand, it is possible to design a system with a predefined upper bound for the bit error rate and derive λ_{opt} and variable watermark capacity determined by block length N.

The developed audio watermarking algorithm has been tested using a large set of songs (Paper X). Both mp3 and low pass filtering attacks have dramatically increased the detection bit error rate, due to the unpredictability of SNR variations, including a complete fade

of the particular frequency subbands, during the watermark data transmission. It is clear that the detection performance of the system using the attack characterization and ISS modulation is significantly higher compared to the method using the standard SS modulation. At lower watermark capacities gains are equal to a few orders of magnitudes in the detection bit error rate (Paper X). However, the bit error rate of the described system was, as expected, still larger than in the case of the ISS modulation system with the non-blind detection. The experimental results have confirmed the algorithm's property to take advantage of the statistical properties of ISS modulation while maintaining a blind detection during the watermark extraction process.

5.7 Summary

The third research subproblem was identified using the following question: How can an overall robustness to the attacks of a watermark system be increased using an attack characterization at the embedding side? Chapter 5 concentrated on increasing the robustness of the embedded watermarks using the attack characterization. Novel principles important for our attack characterization implementation are presented, as well as the watermark channel models of interest.

A particular watermark channel model that was studied was a watermark channel model in the presence of the MPEG compression. We showed that a far more appropriate model for the watermark channel in the presence of mp3 coding is the Rayleigh frequency-selective fading model, because it describes more precisely the distortions that appear. The experimental results suggest that the noise introduced by mp3 compression can hardly be modeled as AWGN and that the BER curves obtained by the Rayleigh fading channel model have steepness and values more close to the experimentally derived ones. The results confirmed that a far better watermark channel modeling is obtained by the proposed model than with the usual AWGN watermark channel model.

Using the available theoretical background, we developed a novel audio watermarking scheme that uses the attack characterization in order to obtain a high robustness against standard watermark attacks. The watermark embedding and detection are based on the frequency hopping spread spectrum method in the spectral domain. The experimental results proved a significant advantage in the detection robustness that the proposed algorithm has, in comparison with the direct sequence spread spectrum algorithm, with a significantly lower BER. In addition, it is clear that the introduction of the attack characterization module additionally improved the extraction reliability of both algorithms, reducing the bit error rate, most discernibly in the presence of MPEG compression and low pass filtering. The overall algorithm obtained a high detection robustness, while decreasing the computational complexity and increasing the perceptual transparency of the watermarked signal.

At the end, it was shown that the attack characterization algorithm that was proposed can be successfully used in other schemes as well. The detection performance of the system using the attack characterization and the ISS modulation is significantly higher compared to the method using the standard SS modulation, uses the statistical properties of ISS modulation while maintaining a blind detection during the watermark extraction.

6 Conclusions

Robust digital audio watermarking algorithms and high capacity steganography methods for audio are studied in this thesis. The main results of this work are the development of novel audio watermarking algorithms, with the state-of-the-art performance and an acceptable increase in the computational complexity. The algorithms' performance is validated in the presence of the standard watermarking attacks. The main point of the "magic triangle" concept is that if the perceptual transparency parameter is fixed, the design of a watermark system cannot obtain a high robustness and watermark data rate at the same time. Therefore, the research problem was divided into three specific subproblems.

Chapter 2 gives an extensive literature review and describes in detail different concepts of watermarking of digital audio. The scientific publications included into the literature survey have been chosen in order to build a sufficient background that would help out in solving the research problems.

The first research subproblem was characterized by the following question: What is the highest watermark bit rate obtainable, under perceptual transparency constraint, and how to approach the limit? The general background and requirements for high bit rate covert communications for audio were given in Chapter 3.

The details and experimental results for the modified time domain LSB steganography algorithm were discussed. The results of subjective tests showed that the perceptual quality of watermarked audio, when embedding is done by the proposed algorithm, is higher in comparison with the standard LSB embedding. The tests confirmed that the described algorithm succeeds in increasing the bit rate of the hidden data for one third without affecting the perceptual transparency of the resulting audio signal. However, the simple LSB coding method in time domain is able to inaudibly embed only 3-4 bits per sample, which is far from a theoretically achievable rate, mostly due to a poor shaping of the noise introduced by embedding and operation in time domain. Therefore, a perceptual entropy and information theoretic assessment of the achievable data rates of a data hiding channel was necessary to develop a scheme that could obtain higher data rates.

A high bit rate algorithm in wavelet domain was developed based on these findings. The wavelet domain was chosen for data hiding due to its low processing noise and suitability for frequency analysis, because of its multiresolutional properties that provide ac-

cess both to the most significant parts and details of signal's spectrum. The experiments showed that the wavelet information hiding scheme has a large advantage over the time domain LSB algorithm. The wavelet domain algorithm produces stego objects perceptually hardly discriminated from the original audio clip even when 8 LSBs of coefficients are modified, providing up to 5 bits per sample higher data rate in comparison with the time domain LSB algorithm.

The second subproblem was defined by the following question: How can the detection performance of a watermarking system be improved using algorithms based on communications models for that system? In Chapter 4, a general model for a spread spectrum-based watermarking is described as well, in order to place in context the developed algorithms.

A spread spectrum audio watermarking algorithm in time domain is presented. The overall watermark detection robustness of the algorithm is comparable to other state-of-the art algorithms, specifically in the presence of mp3 compression, resampling and low pass filtering. On the other hand, the algorithm uses computationally low demanding embedding and detection methods and a simple perceptual model for describing two masking properties of the HAS. One of the malicious attacks on this scheme is the desynchronization of the correlation calculation by time-scale modifications, such as the stretching of the audio sequence or insertion or deletion of samples. In that case, the watermark detection scheme does not properly determine the value of the embedded watermark, resulting in a high increase of the bit error rate.

A resynchronization algorithm that is able to provide a correct watermark detection even in the presence of these attacks, while maintaining a perceptual transparency by a perceptual noise shaping is presented subsequently. The consequence of an improved watermark decoding is a decreased bit rate of the embedded watermark; however the bit rate is still within an acceptable range for most copyright applications.

The possibility of improving the robustness of watermark detection and increasing the resistance to attacks was studied. An audio decorrelation algorithm for the spread-spectrum watermarking that uses least squares Savitzky-Golay smoothing filters is proposed. The test results showed a significant improvement in the detection performance of the described method, compared to the standard watermark detection, especially if a watermarked audio sequence is attacked with mp3 compression or low pass filtering attacks.

In order to further improve the detection robustness and decrease the bit error rate, channel coding was employed, because it has a property to reduce BER for a given watermark bit rate in comparison with the regular detection or equivalently increase an available watermark bit rate for a given BER. The simulations showed that a channel coding maintains a reliable watermark bit rate for a fixed BER, even after severe attacks. However, the introduction of the described turbo channel coding is justified only when the SNR per symbol value is high enough and the iterative decoding of soft output values is able to make the coding gain. One of the implementation issues was the harsh slope of the watermark bit rate vs. BER curve and the sensitivity to cut attack, because the whole block of bits is needed during decoding.

The third subproblem was identified using the following question: How can the overall robustness to the attacks of a watermark system be increased using an attack characterization at the embedding side? Chapter 5 focused on increasing the robustness of the

embedded watermarks using the attack characterization. Novel principles important for our attack characterization implementation are presented, as well as watermark channel models of interest.

The particular watermark channel model that was studied was a watermark channel model in the presence of MPEG compression. We showed that a far more appropriate model for the watermark channel in the presence of mp3 coding is the Rayleigh frequency-selective fading model, because it describes more precisely the distortions that appear. The experimental results suggest that the noise introduced by mp3 compression can hardly be modeled as AWGN and that BER curves obtained by the Rayleigh fading channel model have steepness and values more close to the experimentally derived ones. The results confirmed that a far better watermark channel modeling is obtained by the proposed model than with the usual AWGN watermark channel model.

Using the available theoretical background, we developed a novel audio watermarking scheme that uses the attack characterization in order to obtain a high robustness against standard watermark attacks. The watermark embedding and detection are based on the frequency hopping method in the FFT domain. The experimental results proved a significant advantage in the detection robustness that the proposed algorithm has, in comparison with a direct sequence spread spectrum algorithm, with a significantly lower BER. In addition, it is clear that the introduction of the attack characterization module additionally improved the extraction reliability of both algorithms, decreasing the bit error rate, most discernibly in the presence of MPEG compression and low pass filtering. The overall algorithm obtained a high detection robustness, while decreasing the computational complexity and increasing the perceptual transparency of the watermarked signal.

At the end, it was shown that the attack characterization algorithm that was proposed can be successfully used in other schemes as well. The detection performance of the system using an attack characterization and the ISS modulation is significantly higher compared to the method using the standard SS modulation, uses the statistical properties of ISS modulation while maintaining a blind detection during the watermark extraction.

References

1. Yu H, Kundur D & Lin C (2001) Spies, thieves, and lies: The battle for multimedia in the digital era. IEEE Multimedia 8(3): p 8–12.

2. Cox I, Miller M & Bloom J (2003) Digital Watermarking. Morgan Kaufmann Publishers, San Franciso, CA.

3. Wu M & Liu B (2003) Multimedia Data Hiding. Springer Verlag, New York, NY.

4. Kundur D (2001) Watermarking with diversity: Insights and implications. IEEE Multimedia 8(4): p 46–52.

5. Bloom J, Cox I, Kalker T, Linnartz J, Miller M & Traw C (1999) Copy protection for dvd video. Proceedings of the IEEE 87(7): p 1267–1276.

6. Eggers J & Girod B (2002) Informed Watermarking. Kluwer Academic Publishers, Boston, MA.

7. Johnson N, Duric Z & Jajodia S (2001) Information Hiding: Steganography and Watermarking-Attacks and Countermeasures. Kluwer Academic Publishers, Boston, MA.

8. Anderson R & Petitcolas F (1998) On the limits of steganography. IEEE Journal on Selected Areas in Communications 16(4): p 474–481.

9. Johnson N & Jajodia S (1998) Steganalysis: the investigation of hidden information. In: Proc. IEEE Information Technology Conference, Syracuse, NY, p 113–116.

10. Katzenbeisser S & Petitcolas F (1999) Information Hiding Techniques for Steganography and Digital Watermarking. Artech House, Norwood, MA.

11. Bender W, Gruhl D & Morimoto N (1996) Techniques for data hiding. IBM Systems Journal 35(3): p 313–336.

12. Cox I & Miller M (2001) Electronic watermarking: the first 50 years. In: Proc. IEEE Workshop on Multimedia Signal Processing, Cannes, France, p 225–230.

13. Hartung F & Kutter M (1999) Multimedia watermarking techniques. Proceedings of the IEEE 87(7): p 1709–1107.

14. Swanson M, Zhu B & Tewfik A (1999) Current state-of-the-art, challenges and future directions for audio watermarking. In: Proc. IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, p 19–24.

15. Pan D (1995) A tutorial on mpeg/audio compression. IEEE Multimedia 2(2): p 60–74.

16. Noll P (1993) Wideband speech and audio coding. IEEE Communications Magazine 31(11): p 34–44.

17. Wu M, Trappe W, Wang Z & Liu K (2004) Collusion-resistant fingerprinting for multimedia. IEEE Signal Processing Magazine 21(2): p 15–27.

18. Trappe W, Wu M, Wang Z & Liu K (2003) Anti-collusion fingerprinting for multimedia. IEEE Transactions on Signal Processing 51(4): p 1069–1087.

19. Chenyu W, Jie Z & Zhao-qi B.and Gang R (2003) Robust crease detection in fingerprint images. In: Proc. IEEE International Conference on on Computer Vision and Pattern Recognition, Madison, WI, p 505–510.

20. Hong Z, Wu M, Wang Z & Liu K (2003) Nonlinear collusion attacks on independent fingerprints for multimedia. In: Proc. IEEE Computer Society Conference on Multimedia and Expo, Baltimore, MD, p 613–616.

21. Wang Z, Wu M, Hong Z, Liu K & Trappe W (2003) Resistance of orthogonal gaussian fingerprints to collusion attacks. In: Proc. IEEE Computer Society Conference on Multimedia and Expo, Baltimore, MD, p 617–620.

22. Kirovski D, Malvar H & Yacobi Y (2002) Multimedia content screening using a dual watermarking and fingerprinting system. In: Proc. ACM Multimedia, Juan Les Pins, France, p 372–381.

23. Boneh D & Shaw J (1998) Collusion-secure fingerprinting for digital data. IEEE Transactions on Information Theory 44(9): p 1897–1905.

24. Yacobi Y (2001) Improved boneh-shaw content fingerprinting. In: Proc. Cryptographer's Track at RSA Conference, San Francisco, CA, p 378–391.

25. Dittmann J, Schmitt P, Saar E, Schwenk J & Ueberberg J (2000) Combining digital watermarks and collusion secure fingerprints for digital images. SPIE Journal on Electronic Imaging 9(4): p 456–467.

26. Termont P, De Stycker L, Vandewege J, Op de Beeck M, Haitsma J, Kalker T, Maes M & Depovere G (2000) How to achieve robustness against scaling in a real-time digital watermarking system for broadcast monitoring. In: Proc. IEEE International Conference on Image Processing, Vancouver, BC, p 407–410.

27. Termont P, De Strycker L, Vandewege J, Haitsma J, Kalker T, Maes M, Depovere G, Langell A, Alm C & Norman P (1999) Performance measurements of a real-time digital watermarking system for broadcast monitoring. In: Proc. IEEE International Conference on Multimedia Computing and Systems, Florence, Italy, p 220–224.

28. Depovere G, Kalker T, Haitsma J, Maes M, de Strycker L, Termont P, Vandewege J, Langell A, Alm C, Norman P, O'Reilly G, Howes B, Vaanholt H, Hintzen R, Donnelly P & Hudson A (1999) The viva project: digital watermarking for broadcast monitoring. In: Proc. IEEE International Conference on Image Processing, Kobe, Japan, p 202–205.

29. Kalker T & Haitsma J (2000) Efficient detection of a spatial spread-spectrum watermark in mpeg video streams. In: Proc. IEEE International Conference on Image Processing, Vancouver, BC, p 407–410.

30. Craver S & Stern J (2001) Lessons learned from sdmi. In: Proc. IEEE International Workshop on Multimedia Signal Processing, Cannes, France, p 213–218.

31. Arnold M, Wolthusen S & Schmucker M (2003) Techniques and Applications of Digital Watermarking and Content Protection. Artech House, Norwood, MA.

32. Johnston J (1998) Estimation of perceptual entropy using noise masking criteria. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, New York, NY, p 2524–2527.

33. Kirovski D & Malvar H (2001) Spread-spectrum audio watermarking: Requirements, applications, and limitations. In: Proc. IEEE International Workshop on Multimedia Signal Processing, Cannes, France, p 219–224.

34. Zwicker E & Fastl H (1999) Psychoacoustics. Springer Verlag, Berlin, Germany.

35. Schneier B (1996) Applied Cryptography. John Wiley and Sons, Indianapolis, IN.

36. Mallat S (2001) Wavelet Tour of Signal Processing. Academic Press, San Diego, CA.

37. Cover T & Thomas J (1991) Elements of Information Theory. John Wiley and Sons, Indianapolis, IN.

38. Chen B & Wornell G (1999) Dither modulation: a new approach to digital watermarking and information embedding. In: Proc. SPIE: Security and Watermarking of Multimedia Contents, San Hose, CA, p 342–353.

39. Cox I, Miller M & McKellips A (1999) Watermarking as communications with side information. Proceedings of the IEEE 87(7): p 1127 –1141.

40. Shannon C (1958) Channels with side information at the transmitter. IBM Journal of Research and Development 2(7): p 289–293.

41. Costa H (1983) Writing on dirty paper. IEEE Transactions on Information Theory 9(3): p 439–441.

42. Perez-Gonzalez F & Balado F (2002) Quantized projection data hiding. In: Proc. IEEE International Conference on Image Processing, Rochester, NY, p 889–892.

43. Eggers J, Bäuml R, Tzschoppe R & Girod B (2003) Scalar costa scheme for information embedding. IEEE Transactions on Signal Processing 51(4): p 1003–1019.

44. Su J & Girod B (2002) Power-spectrum condition for energy-efficient watermarking. IEEE Transactions on Multimedia 4(4): p 551–560.

45. Su J, Eggers J & Girod B (2001) Analysis of digital watermarks subjected to optimum linear filtering and additive noise. Signal Processing 81(6): p 1141–1175.

46. Moulin P & O'Sullivan J (2003) Information-theoretic analysis of information hiding. IEEE Transactions on Information Theory 49(3): p 563–593.

47. Moulin P (2001) The role of information theory in watermarking and its application to image watermarking. Signal Processing 81(6): p 1121–1139.

48. Fridrich J, Goljan M & Du R (2001) Distortion-free data embedding. Lecture Notes in Computer Science 2173: p 27–41.

49. Lee Y & Chen L (2000) High capacity image steganographic model. IEE Proceedings Vision Image Signal Processing 147(3): p 288–294.

50. Fridrich J, Goljan M & Du R (2002) Lossless data embedding - new paradigm in digital watermarking. Applied Signal Processing 2002(2): p 185–196.

51. Yeh C & Kuo C (1999) Digital watermarking through quasi m-arrays. In: Proc. IEEE Workshop on Signal Processing Systems, Taipei, Taiwan, p 456–461.

52. Cedric T, Adi R & Mcloughlin I (2000) Data concealment in audio using a nonlinear frequency distribution of prbs coded data and frequency-domain lsb insertion. In: Proc. IEEE Region 10 International Conference on Electrical and Electronic Technology, Kuala Lumpur, Malaysia, p 275–278.

53. Mobasseri B (1998) Direct sequence watermarking of digital video using m-frames. In: Proc. IEEE International Conference on Image Processing, Chicago, IL, p 399–403.

54. Chandramouli R & Memon N (2001) Analysis of lsb based image steganography techniques. In: Proc. IEEE International Conference on Image Processing, Thessaloniki, Greece, p 1019–1022.

55. Fridrich J, Goljan M & Du R (2001) Detecting lsb steganography in color, and gray-scale images. IEEE Multimedia 8(4): p 22–28.

56. Dumitrescu S, Wu X & Wang Z (2003) Detection of lsb steganography via sample pair analysis. IEEE Transactions on Signal Processing 51(7): p 1995–2007.

57. Ruiz F & Deller J (2000) Digital watermarking of speech signals for the national gallery of the spoken word. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, p 1499–1502.

58. Ciloglu T & Karaaslan S (2000) An improved all-pass watermarking scheme for speech and audio. In: Proc. IEEE International Conference on Multimedia and Expo, New York, NY, p 1017–1020.

59. Tilki J & Beex A (1997) Encoding a hidden auxiliary channel onto a digital audio signal using psychoacoustic masking. In: Proc. IEEE South East Conference, Blacksburg, VA, p 331–333.

60. Lancini R, Mapelli F & Tubaro S (2002) Embedding indexing information in audio signal using watermarking technique. In: Proc. 4th EURASIP-IEEE International Symposium on Video/Image Processing and Multimedia Communications, Zadar, Croatia, p 257–261.

61. Kuo S, Johnston J, Turin W & Quackenbush S (2002) Covert audio watermarking using perceptually tuned signal independent multiband phase modulation. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, p 1753–1756.

62. Gang L, Akansu A & Ramkumar M (2001) Mp3 resistant oblivious steganography. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT, p 1365–1368.

63. Huang D & Yeo T (2002) Robust and inaudible multi-echo audio watermarking. In: Proc. IEEE Pacific-Rim Conference On Multimedia, Taipei, China, p 615–622.

64. Ko B, Nishimura R & Suzuki Y (2002) Time-spread echo method for digital audio watermarking using pn sequences. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, p 2001–2004.

65. Foo S, Yeo T & Huang D (2001) An adaptive audio watermarking system. In: Proc. IEEE Region 10 International Conference on Electrical and Electronic Technology, Phuket Island-Langkawi Island, Singapore, p 509–513.

66. Xu C, Wu J, Sun Q & Xin K (1999) Applications of watermarking technology in audio signals. Journal Audio Engineering Society 47(10): p 1995–2007.

67. Oh H.O. Seok J, Hong J & Youn D (2001) New echo embedding technique for robust and imperceptible audio watermarking. In: IEEE International Conference on Acoustic, Speech and Signal Processing, Salt Lake City, UT, p 1341–1344.

68. Bassia P, Pitas I & Nikolaidis N (2001) Robust audio watermarking in the time domain. IEEE Transactions on Multimedia 3(2): p 232–241.

69. Neubauer C, Herre J & Brandenburg K (1998) Continuous steganographic data transmission using uncompressed audio. In: Proc. Information Hiding Workshop, Portland, OR, p 208–217.

70. Cox I, Kilian J, Leighton F & Shamoon T (1997) Secure spread spectrum watermarking for multimedia. IEEE Transactions on Image Processing 6(12): p 1673–1687.

71. Kirovski D & Malvar H (2003) Spread-spectrum watermarking of audio signals. IEEE Transactions on Signal Processing 51(4): p 1020–1033.

72. Swanson M, Zhu B, Tewfik A & Boney L (1998) Robust audio watermarking using perceptual masking. Signal Processing 66(3): p 337–355.

73. Neubauer C & Herre J (2000) Advanced audio watermarking and its applications. In: Proc. AES Convention, Audio Engineering Society preprint 5176, Los Angeles, CA, p 311–319.

74. Neubauer C & Herre J (1998) Digital watermarking and its influence on audio quality. In: Proc. AES Convention, Audio Engineering Society preprint 4823, San Francisco, CA, p 225–233.

75. Ikeda M, Takeda K & Itakura F (1999) Audio data hiding use of band-limited random sequences. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Phoenix, AZ, p 2315–2318.

76. Seok J & Hong J (2001) Audio watermarking for copyright protection of digital audio data. Electronics Letters 37(1): p 60–61.

77. Kalker T & Janssen A (1999) Analysis of watermark detection using spomf. In: Proc. IEEE International Conference on Image Processing, Kobe, Japan, p 889–892.

78. Kirovski D & Malvar H (2001) Robust covert communication over a public audio channel using spread spectrum. In: Proc. Information Hiding Workshop, Pittsburgh, PA, p 889–899.

79. Saito S, Furukawa T & Konishi K (2002) A digital watermarking for audio data using band division based on qmf bank. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, p 3473–3476.

80. Tachibana R, Shimizu S, Kobayashi S & Nakamura T (2002) An audio watermarking method using a two-dimensional pseudo-random array. Signal Processing 82(10): p 1455–1469.

81. Li X & Yu H (2000) Transparent and robust audio data hiding in subband domain. In: Proc. International Conference on Information Technology: Coding and Computing, Las Vegas, NV, p 74–79.

82. Lee S & Ho Y (2000) Digital audio watermarking in the cepstrum domain. Electronics Letters 46(3): p 744–750.

83. Li X & Yu H (2000) Transparent and robust audio data hiding in cepstrum domain. In: Proc. IEEE International Conference on Multimedia and Expo, New York, NY, p 397–400.

84. Neubauer C & Herre J (2000) Audio watermarking of mpeg-2 aac bitstreams. In: Proc. AES Convention, Audio Engineering Society preprint 5101, Paris, France, p 395–404.

85. Cheng S, Yu H & Xiong Z (2002) Enhanced spread spectrum watermarking of mpeg-2 aac. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, p 3728–3731.

86. Furon T & Duhamel P (2003) An asymmetric watermarking method. IEEE Transactions on Signal Processing 51(4): p 981–995.

87. Tefas A, Nikolaidis A, Nikolaidis N, Solachidis V, Tsekeridou S & Pitas I (2003) Performance analysis of correlation-based watermarking schemes employing markov chaotic sequences. IEEE Transactions on Signal Processing 51(4): p 1979–1994.

88. Barni M, Bartolini F, De Rosa A & Piva A (2003) Optimum decoding and detection of multiplicative watermarks. IEEE Transactions on Signal Processing 51(4): p 1118–1123.

89. Lipschutz S & Lipson M (2000) Schaum's Outline of Linear Algebra. McGraw-Hill, New York, NY.

90. Miller M, Cox I & Bloom J (2000) Informed embedding: Exploiting image and detector information during watermark insertion. In: Proc. IEEE International Image Processing Conference, Vancouver, BC, p 1–4.

91. Malvar H & Florencio D (2003) Improved spread spectrum: A new modulation technique for robust watermarking. IEEE Transactions on Signal Processing 51(4): p 898–905.

92. Sugihara R (2001) Practical capacity of digital watermark as constrained by reliability. In: Proc. International Conference on Information Technology: Coding and Computing, Las Vegas, NV, p 85–89.

93. Arnold M (2000) Audio watermarking: features, applications and algorithms. In: Proc. IEEE International Conference on Multimedia and Expo, New York, NY, p 1013–1016.

94. Yeo I & Kim H (2003) Modified patchwork algorithm: A novel audio watermarking scheme. IEEE Transactions on Speech and Audio Processing 11(4): p 381–386.

95. Arnold M & Huang Z (2001) Blind detection of multiple audio watermarks. In: Proc. International Conference on Web Delivering of Music, Florence, Italy, p 12–19.

96. Xu C, Wu J & Sun Q (1999) A robust digital audio watermarking technique. In: Proc. International Symposium on Signal Processing and its Applications, Brisbane, Australia, p 95–98.

97. Xu C, Wu J & Sun Q (1999) Digital audio watermarking and its application in a multimedia database. In: Proc. International Symposium on Signal Processing and its Applications, Brisbane, Australia, p 91–94.

98. Lemma A, Aprea J, Oomen W & Van de Kerkhof L (2003) A temporal domain audio watermarking technique. IEEE Transactions on Signal Processing 51(4): p 1088–1097.

99. Kaabneh K & Youssef A (2001) Muteness-based audio watermarking technique. In: Proc. International Conference on Distributed Computing Systemss, Phoenix, AZ, p 379–383.

100. Yang H, Patra J & Chan C (2002) An artificial neural network-based scheme for robust watermarking of audio signals. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Orlando, FL, p 1029–1032.

101. Hsieh C & Sou P (2002) Blind cepstrum domain audio watermarking based on time energy features. In: Proc. 14th International Conference on Digital Signal Processing, Santorini, Greece, p 705–708.

102. Mansour M & Tewfik A (2001) Audio watermarking by time-scale modification. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT, p 1353–1356.

103. Xu C & Feng D (2002) Robust and efficient content-based digital audio watermarking. Multimedia Systems 8(5): p 353–368.

104. Lie W & Chang L (2001) Robust and high-quality time-domain audio watermarking subject to psychoacoustic masking. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Sydney, Australia, p 45–48.

105. Chou J, Ramchandran K & Ortega A (2001) High capacity audio data hiding for noisy channels. In: Proc. International Conference on Information Technology: Coding and Computing, Las Vegas, NV, p 108–111.

106. Servetto S, Podilchuk C & Ramchandran K (1998) Capacity issues in digital image watermarking. In: Proc. IEEE International Conference on Image Processing, Chicago, IL, p 445–449.

107. Chou J, Pradhan S, El Ghaoui L & Ramchandran K (2000) A robust optimization solution to the data hiding problem using distributed source coding principles. In: Proc. of SPIE: Security and Watermarking of Multimedia Contents, San Jose, CA, p 325–339.

108. Chou J, Pradhan S, El Ghaoui L & Ramchandran K (2000) Watermarking based on duality with distributed source coding techniques and robust optimization principles. In: Proc. IEEE International Conference on Image Processing, Vancouver, BC, p 585–588.

109. Pradhan S, Chou J & Ramchandran K (2003) Duality between source coding and channel coding and its extension to the side information case. IEEE Transactions on Information Theory 49(5): p 1181–1203.

110. Hartung F & Girod B (1998) Watermarking of uncompressed and compressed video. Signal Processing 66(3): p 283–301.

111. Cohen A & Lapidoth A (2000) On the gaussian watermarking game. In: Proc. IEEE International Symposium on Information Theory, Sorrento, Italy, p 48.

112. Jabri A & Albawardi W (2000) Characterization of digital images as a communication channel for steganographic applications. In: Proc. Canadian Conference on Electrical and Computer Engineering, Toronto, ON, p 114–117.

113. Katsavounidis I & Jay Kuo C (1997) A multiscale error diffusion technique for digital halftoning. IEEE Transactions on Image Processing 6(3): p 1181–1203.

114. Mintzer F, Goertzil G & Thompson G (1992) Display of images with calibrated color on a system featuring monitors with limited color palettes. In: Digest of technical Papers SID International Symposium, Boston, MA, p 377–380.

115. Yeung M & Mintzer F (1997) An invisible watermarking technique for image verification. In: Proc. IEEE International Conference on Image Processing, Washington, DC, p 680–683.

116. Johnston J (1988) Transform coding of audio signals using perceptual noise criteria. IEEE Journal on Selected Areas in Communications 6(2): p 314–323.

117. Ramkumar M & Akansu A (1988) Information theoretic bounds for data hiding in compressed images. In: Proc. IEEE Workshop on Multimedia Signal Processing, Los Angeles, CA, p 267–272.

118. Ramkumar M & Akansu A (2001) Capacity estimates for data hiding in compressed images. IEEE Journal on Selected Areas in Communications 10(8): p 1252–1263.

119. Boney L, Tewfik A & Hamdy K (1996) Digital watermarks for audio signals. In: Proc. IEEE International Conference on Multimedia Computing and Systems, Hiroshima, Japan, p 473–480.

120. Gordy J & L. B (2000) Performance evaluation of digital audio watermarking algorithms. In: Proc. IEEE Midwest Symposium on Circuits and Systems, Michigan State University, MI, p 456–459.

121. Laftsidis C, Tefas A, Nikolaidis N & Pitas I (2003) Robust multibit audio watermarking in the temporal domain. In: Proc. International Symposium on Circuits and Systems, Bangkok, Thailand, p 944–947.

122. Cvejic N & Seppanen T (2003) Robust audio watermarking in wavelet domain using frequency hopping and modified patchwork method. In: Proc. International Symposium on Image and Signal Processing and Analysis, Rome, Italy, p 251–255.

123. Petitcolas F (2000) Watermarking schemes evaluation. IEEE Signal Processing Magazine 17(5): p 58–64.

124. Eskicioglua A, Townb J & Delp E (2003) Security of digital entertainment content from creation to consumption. IEEE Signal Processing Magazine 18(4): p 237–262.

125. Steinebach M, Petitcolas F, Raynal F, Dittmann J, Fontaine C, Seibel S, Fates N & Ferri L (2001) Stirmark benchmark: Audio watermarking attacks. In: Proc. International Conference on Information Technology: Coding and Computing, Las Vegas, NV, p 49–54.

126. Voloshynovski S, Pereira S, Iquise V & Pun T (2001) Attack modelling: towards a second generation watermarking benchmark. Signal Processing 81(6): p 1177–1214.

127. Miller M, Dorr G & Cox I (2002) Dirty-paper trellis codes for watermarking. In: Proc. IEEE International Conference on Image Processing, Rochester, NY, p 129–132.

128. Hernandez J & Perez-Gonzalez F (1999) Statistical analysis of watermarking schemes for copyright protection of images. Proceedings of the IEEE 87(7): p 1142–1166.

129. Kundur D & Hatzinakos D (2001) Diversity and attack characterization for improved robust watermarking. IEEE Transactions on Signal Processing 29(10): p 2383–2396.

130. Barni M, Bartolini F, De Rosa A & Piva A (2003) Optimum decoding and detection of multiplicative watermarks. IEEE Transactions on Signal Processing 51(4): p 1118–1123.

131. Perez-Gonzalez F, Hernandez J & Balado F (2001) Approaching the capacity limit in image watermarking: A perspective on coding techniques for data hiding applications. Signal Processing 81(6): p 1215–1238.

132. Hernandez J, Rodriguez J & Perez-Gonzalez F (2001) Improving the performance of spatial watermarking of images using channel coding. Signal Processing 80(7): p 1261–1279.

133. Comesana P, Perez-Gonzalez F & Balado F (2003) Optimal strategies for spread-spectrum and quantized-projection image data hiding games with ber payoffs. In: Proc. IEEE International Conference on Image Processing, Barcelona, Spain, p 145–148.

134. Miller M & Bloom J (1999) Computing the probability of false watermark detection. In: Proc. Workshop on Information Hiding, Dresden, Germany, p 49–54.

135. Jiang D, Weixin X & Jianping Y (2000) Study on capacity of information hiding for still image. In: Proc. International Conference on Signal Processing, Beijing, China, p 1010 – 1013.

136. Cox I & Miller M (2002) Preprocessing media to facilitate later insertion of a watermark. In: Proc. International Conference on Digital Signal Processing, Santorini, Greece, p 67–70.

137. Briassouli A & Moulin P (2003) Detection-theoretic anaysis of warping attacks in spread-spectrum watermarking. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, p 53–56.

138. Orfanidis S (1996) Introduction to Signal Processing. Prentice-Hall, Englewood Cliffs, NJ.

139. Cromwell J, Labys W & Terraza M (1994) Univariate Tests for Time Series Models. Sage, Thousand Oaks, CA.

140. Conover W (1980) Practical Nonparametric Statistics. John Wiley and Sons, New York, NY.

141. Liu T & Moulin P (2003) Error exponents for one-bit watermarking. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Hong Kong, China, p 65–68.

142. Liu T & Moulin P (2003) Error exponents for watermarking game with squared-error constraints. In: Proc. International Symposium on Information Theory, Yokohama, Japan, p 190.

143. Moulin P & Ivanovic A (2003) The zero-rate spread-spectrum watermarking game. IEEE Transactions on Signal Processing 51(4): p 1098–1117.

144. Karakos D & Papamarcou A (2003) A relationship between quantization and watermarking rates in the presence of additive gaussian attacks. IEEE Transactions on Information Theory 49(8): p 1970–1982.

145. Depovere G, Kalker T & Linnartz J (1998) Improved watermark detection reliability using filtering before correlation. In: Proc. IEEE International Conference on Image Processing, Chicago, IL, p 430–434.

146. Poor H (1994) An Introduction to Signal Detection and Estimation. Springer Verlag, New York, NY.

147. Vucetic B & Juan Y (2000) Turbo Codes: Principles and Applications. Kluwer Academic Publishers, Boston, MA.

148. Ambroze A, Wade G, Serdean C, Tomlinson M, Stander J & Borda M (2001) Turbo code protection of video watermark channel. IEE Proceedings Vision Image Signal Processing 148(1): p 54–58.

149. Perez-Gonzalez F & Balado F (2001) Coding at the sample level for data hiding: Turbo and concatenated codes. In: Proc. of SPIE: Security and Watermarking of Multimedia Contents, San Jose, CA, p 532–543.

150. Balado F, Perez-Gonzalez F & Scalise S (2001) Turbo coding for sample-level watermarking in the dct domain. In: Proc. IEEE International Conference on Image Processing, Thessalonica, Greece, p 1003–1006.

151. Kesal M, Mihcak M, Koetter R & Moulin P (2000) Iteratively decodable codes for watermarking applications. In: Proc. International Symposium on Turbo Codes and Related Topics, Brest, France, p 589–596.

152. Loo P & Kingsbury N (2002) Watermark detection based on the properties of error control codes. IEE Proceedings Vision Image Signal Processing 150(2): p 115–121.

153. Baudry S, Delaigle J, Sankur B, Macq B & Matre H (2002) Analyses of error correction strategies for typical communication channels in watermarking. Signal Processing 81(6): p 1239–1250.

154. Baudry S, Nguyen P & Maitre H (2000) Channel coding in video watermarking: use of soft decoding to improve the watermark retrieval. In: Proc. IEEE International Conference on Image Processing, Vancouver, BC, p 25–28.

155. Gu L, Huang J & Shi Y (2003) Analysis of the role played by error correcting coding in robust watermarking. In: Proc. International Symposium on Circuits and Systems, Bangkok, Thailand, p 798–801.

156. Furon T, Moreau N & Duhamel P (2000) Audio public key watermarking technique. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Istanbul, Turkey, p 1959–1962.

157. Dittmann J, Mukherjee A & Steinbach M (2000) Media-independent watermarking classification and need for combining digital video and audio watermarking for media authentication. In: Proc. International Conference on Information Technology, Las Vegas, NV, p 62–67.

158. Lee S & Jung S (2001) A survey of watermarking techniques applied to multimedia. In: Proc. IEEE International Symposium on Industrial Electronics, Pusan, Korea, p 272–277.

159. Campisi P, Carli M, Giunta G & Neri A (2003) Blind quality assessment system for multimedia communications using tracing watermarking. IEEE Transactions on Signal Processing 51(4): p 996–1002.

160. Zhang X & Wang S (2002) Watermarking scheme capable of resisting attacks based on availability of inserter. Signal Processing 82(11): p 1801–1804.

161. Su K, Kundur D & Hatzinakos D (2001) A content-dependent spatially localized video watermarked for resistance to collusion and interpolation attacks. In: Proc. IEEE International Conference on Image Processing, Thessalonica, Greece, p 818–821.

162. Kundur D & Hatzinakos D (1999) Digital watermarking for telltale tamper-proofing and authentication. Proceedings of the IEEE 87(7): p 1167–1180.

163. Wu M & Liu B (2003) Data hiding in image and video .i. fundamental issues and solutions. IEEE Transactions on Image Processing 12(6): p 685–695.

164. Wu M & Liu B (2003) Data hiding in image and video .ii. designs and applications. IEEE Transactions on Image Processing 12(6): p 696–705.

165. Burgett S, Koch E & Zhao J (1998) Copyright labeling of digitized image data. IEEE Communications Magazine 36(3): p 94–100.