Audio Information Hiding Based on Singular-Spectrum Analysis with Psychoacoustic Model and Its Applications

Jessada KARNJANA

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

Audio Information Hiding Based on Singular-Spectrum Analysis with Psychoacoustic Model and Its Applications

Jessada KARNJANA

Supervisor: Associate Professor Masashi UNOKI

School of Information Science
Japan Advanced Institute of Science and Technology

September 2016

Abstract

The growth of the Internet since the last century and the spread of digital-multimedia transfer are useful and convenient for us because it has enabled us to access gigantic shared data. As a consequence, demands for applications such as broadcast monitoring, owner identification, proof of ownership, transaction tracking, tampering detection, copy control, and information carrier for audio signal have increased considerably due to technologies misusage. To answer such demands, audio information hiding has been suggested. There are five requirements for audio information hiding. (1) *Inaudibility*: a property that hidden information does not affect a perceptual quality of host signals. (2) *Robustness*: an ability to extract hidden information correctly when attacks are performed. (3) *Blindness*: a property of extracting hidden information correctly without the original signal. (4) *Confidentiality*: a property of concealing the hidden data. (5) *Capacity*: quantity of hidden information. To meet the first requirement is a real challenge because the human auditory system is very sensitive. When the first and the second are required together, the challenge is tougher because they conflict with each other. Compromising them has proved to be difficult. Actually, not just two, but all requirements conflict with each other.

The aim of this research is to explore audio information hiding that can satisfy all requirements, especially the conflict between inaudibility and robustness. A literature review of various audio-information-hiding techniques has suggested that audio watermarking based on singular value decomposition (SVD) is one of the robust techniques, and the published results are promising. Fundamentally, its robustness is due to the fact that a singular value is invariant under common signal processing. The hidden information is embedded by slightly modifying singular values. However, there are two critical problems. First, the problem about the balance between inaudibility and robustness. All SVD-based schemes treat an audio signal as a meaningless time-series and rely only on a mathematical singular-value-manipulation. They have never taken audio features or human perception into account. Second, when we see them from the acoustic-signal-processing point of view, the physical meaning of singular values has never been addressed. Thus, it seems impossible to formulate a modification rule associating with human perception. The sole philosophy behind the published embedding rules seems to be that, notwithstanding the physical meaning of singular values is unknown, a human being cannot perceive the difference between original and watermarked sounds if the modification is done slightly.

Inspired by these facts, we propose a framework based on the singular-spectrum analysis (SSA), which is closely related to the SVD. We show that, by using SSA, singular values can have the physical meaning. Actually, they are scale factors of oscillatory components of the signal. Hence, by adopting SSA, we can exploit the advantages of SVD-based techniques, and, at the same time, SSA provides us the framework in which a modification rule can be informed. Quite contrary to the philosophy of conventional SVD-based schemes, the philosophy of this work is based on an idea that the embedding rules should be based on both the nature of the audio signal and the human audio-perceptual ability. When combining human perception model, such as the psychoacoustic models, and the strength of the SVD-based technique together, it is expected that the problem of conflicting requirements can be solved. Solving this problem is the ultimate goal.

In this work, we investigate the potentiality of SSA and formulate some basic principles that can be used to achieve the goal. Six audio-information-hiding models based on SSA are proposed. The test results show that the proposed framework achieve five subgoals. The scheme we implement can keep the advantages of the SVD-based technique and, at the same time, can reach a better performance with

the help of an artificial intelligence technique. We also found the connection between singular-spectrum index and peak frequency of oscillatory components and used the finding to improve performance further. In addition, the self-synchronization for watermark detection is proposed. To demonstrate that the framework is practicable, we applied it to applications of ownership protection, information carrier, and fragile audio-watermarking.

Keywords: audio information hiding, singular-spectrum analysis, differential evolution, psychoacoustic model, self-synchronization

Acknowledgments

Doing research is like investigating a crime scene, once said my Sensei, my supervisor Prof. Masashi Unoki. *We researchers are like detectives who collect and analyze data, and then use them to serve our purpose*, he taught. Unlike fictional, brilliant and famous detectives (Doyle's Holmes, Christie's Poirot, Simenon's Maigret, Aoyama's Conan, and you name it), a real detective in science has hardly, if not never, solved a case alone. Many people, too numerous to mention, have helped bring this dissertation to fruition.

My deepest gratitude goes first and foremost to my adviser, Unoki-sensei, for his tremendous guidance and support. I appreciate his broad and deep knowledge and patience every time we discuss. Without his supervision, there seems to be no end in sight to my Ph.D. course. I would like to extend my heartfelt gratitude to my vice supervisor Prof. Masato Akagi whose a lot of insightful questions and comments on my laboratory-meeting reports and presentation rehearsals have helped me to improve my knowledge and skills. Tons of critical comments and suggestions have been from my two Thai professors and co-advisors, Prof. Pakinee Aimmanee and Dr. Chai Wutiwiwatchai, as well. I would like to take this opportunity to thank them for giving me their time and support. I would also like to extend my sincere thanks to Dr. Kwan Sitathani, then the deputy director at NECTEC, who introduced and persuaded me to the SIIT-JAIST-NECTEC dual degree program. Without him (and his call), I might not sit here. Thanks also go to my friends and members of Akagi & Unoki laboratories. They are of great importance for me. I appreciate the supporting grants in SIIT-JAIST-NECTEC dual degree program, A3 Foresight program, and JAIST research grant.

Last but not least, I would like to give many thanks to my parents and two sisters who have stood by my side during almost forty years of my life.

Table of Contents

This dissertation was prepared according to the curriculum for the Collaborative Education Program organized by Japan Advanced Institute of Science and Technology and Sirindhron International Institute of Science and Technology, Thammasat University.

List of Figures

List of Tables

Notation

γ	SMR level
δ	scan step size
Δ	frame-scan step size
ϵ	embedding parameter ϵ
η	overlap degree
$\Lambda_i, \sqrt{\lambda_i}$	singular value at index i
λ_i	eigenvalue at index i
Σ	overlap margin
b_i	concave/convex index
$D_{m,n}$	concavity density of the singular values from indices m to n
k	number of subframes (number of repetitions)
l	lower-bound index (embedding parameter l)
L	window length
$L(i)$	function defining the line connecting Λ_m and Λ_n
N	frame/subframe size
u	upper-bound index (embedding parameter u)
$P_{\mathrm{NM}}(\bar{k})$	noise masker at the geometric mean spectral line $\bar{k}$ of the critical band
$P_{\mathrm{TM}}(k)$	tonal masker at the spectral line k
S_T	tonal component set
$SF(i,j)$	spread of masking from the masker bin j to the maskee bin i
$T_g(i)$	global masking threshold at the frequency bin i
$T_{\mathrm{NM}}(i,j)$	masking contribution at the frequency bin i due to the noise masker located at the bin j
$T_{\mathrm{TM}}(i,j)$	masking contribution at the frequency bin i due to the tonal masker located at the bin j
$T_q(i)$	absolute threshold of hearing at the frequency bin i
$z_b(i)$	Bark frequency of the frequency bin i

Acronym and Abbreviation

AIH	audio information hiding
BDR	bit-detection rate
BER	bit-error rate
bps	bit per second
DE	differential evolution
DWT	discrete wavelet transform
FFT	fast Fourier transform
LSD	log-spectral distance
NMN	noise-masking-noise
NMT	noise-masking-tone
ODG	objective difference grade
PEAQ	perceptual evaluation of audio quality
PSD	power spectral density
RSA	Rivest-Shamir-Adleman
SDR	signal-to-distortion ratio
SMR	signal-to-mask ratio
SPL	sound pressure level
SSA	singular-spectrum analysis
SVD	singular value decomposition
TMN	tone-masking-noise
TMT	tone-masking-tone

Chapter 1

Introduction

1.1 Importance of research and its challenges

Like a coin, the Internet has two sides. The Internet is a good distribution system, thus for the music industry, it means a lot of benefits in terms of market expansion, but, at the same time, it brings the danger of piracy of intellectual property rights [1]. Digital goods have the distinctive characteristic, i.e., they are expensive to produce for the first copy but cheap to reproduce for subsequent [2]. When combining with the great benefits from the digital age, e.g., copying without loss of quality [3], the availability of efficient data compression [4], and the availability of high-speed network [5], the music industry and its artists have become victims [6]. Music sharing via the Internet has been estimated to result in annual sale losses of 3.1 billion US dollars by 2005 [7]. The music industry has been aware of this danger and has sought for a solution since 1990s [1].

The first technology that the industry or contents owners turn to is cryptography [8,9]. Although it can solve many problems concerning confidentiality, data integrity, and authentication [10], there are some problems that cannot be solved by cryptography. For example, once information is decrypted, it is no longer protected. How can cryptography protect the information which is decrypted legally but distributed illegally?

Another problem of cryptography is that it messes up information. It prevents information usage. Actually, downloading digital products for free also has a positive effect due to sampling, i.e., the match between product and customer's tastes is increased [11]. Can cryptography protect information without messing it up? Another example is a

story told by Toby Sharp [12] about a computer engineer who went to work within a restricted country and the encrypted message is restricted. The bottom line is that, in the end, cryptography cannot conceal the fact that two parties communicate to each other secretly.

Alternatively, audio information hiding can be one of potential solutions [13,14]. It is a scheme of making information unnoticeable. In other words, a user is not even aware of the existence of hidden information. Therefore, it can serve the purpose of both information protection and secret communication. In addition, it can protect content after the content is decrypted.

Even though the use of audio information hiding for copyright control has been considered to meet the original goal [15], there are many other types of applications (Examples and details will be given in Chapter 2.).

In general, there are five requirements for audio information hiding [13, 16, 17].

1. **Inaudibility**: a property that hidden information does not affect a perceptual quality of the host signal.
2. **Robustness**: an ability to extract hidden information correctly when attacks are performed.
3. **Blindness**: a property of extracting hidden information correctly without the original signal.
4. **Confidentiality**: a property of concealing hidden data.
5. **Capacity**: quantity of the information embedded into the host signal.

Naturally, these requirements conflict with each other. The high robustness, for example, normally comes with the cost of low audio quality or semi-transparency [18]. Some techniques are good at transparency or inaudibility, but not blind [19]. High capacity implies low robustness [20]. Therefore, in addition to proposing a new effective technique, many researchers in this field have focused on how to compromise these conflicts, and it has proved to be difficult. Therefore, there is a trade-off among these requirements, and to solve the problem of conflicting requirements is one of the challenges in the audio information hiding.

The other challenge comes from the fact that the human auditory system is very sensitive. We can hear a sound wave with extremely small pressure fluctuations [21], and,

by nature, a watermark is nothing but noise added to a host signal. How to fool our ears is a difficult task by itself.

Therefore, from the viewpoints of social concerns and of the scientific and engineering challenges, the research in audio information hiding is of interest.

As a final remark, like the Internet (and a coin), information hiding also has two sides. The hidden information can be used for good or for bad. As reported by the Guardian on Tuesday 7 November 2006 [22], an Al Qaeda operative, Dhiren Barot, concealed his reconnaissance in New York within a copy of Bruce Willis movie *Die Hard* 3. From this view point, the important of information hiding is in analyzing cover signals in order to detect hidden message as well.

1.2 Motivation and research goal

A literature review of many audio-information-hiding techniques has suggested that audio watermarking based on singular value decomposition (SVD) is one of the robust techniques [19, 23]. Fundamentally, in SVD-based information hiding, a watermark bit is embedded by modifying singular values of a matrix representing a host signal, and its robustness is due to the invariance of singular values under common signal processing [24]. However, SVD-based methods have two critical problems.

1. A balance between inaudibility and robustness for certain pieces of music is not good enough, i.e., high inaudibility comes with the cost of sound quality. One of the reasons is that all SVD-based schemes treat an audio signal as a meaningless time-series. They are input-independent and rely only on a mathematical method of extracting singular values. All SVD-based methods have never taken any audio features nor human perception into account [18, 23, 25–43]. In other words, modification rules employed by those schemes are uninformed.

2. When we look from the acoustic-signal-processing point of view, a physical meaning of singular values has never been addressed [18, 23, 25–43]. Thus, it seems impossible to formulate a modification rule associating with human perception when the relation between singular value and physical feature is not established.

The motivation for this research has started from curiosity of what happen when a mathematical manipulation method combines with a human perception model. Is it possible to integrate into each other? Is it possible to bring the advantages of the SVD-based scheme and those of the human perception model together? If all the answers are yes, it is expected that the critical problems stated above can be overcome, and this combination might pave the way to a solution that solves the conflict in requirements. This research aims to explore audio information hiding that can satisfy all requirements, especially the conflict between inaudibility and robustness.

We adopt the singular-spectrum analysis (SSA), which is an SVD-based analysis technique, as a core structure. We choose SSA because, when a signal is analyzed, singular values can be interpreted and have the physical meaning. The physical meaning is needed in order to link SSA to the perceptual model. We use SSA to decompose a signal into a finite number of additive oscillatory components, and singular values are scale factors of their associated components. Then, a watermark bit is embedded by slightly changing scale factors of some oscillatory components. Hence, by adopting SSA, we can exploit the advantages of SVD-based technique, and, at the same time, SSA provides us the framework of which a modification rule which can be informed by human auditory-perception conditions. The central philosophy of this research is that SSA equipped with such perceptual conditions can give a good balance between inaudibility and robustness so that it can overcome the critical problems in SVD-based methods.

The ultimate goal of this work is to resolve the problem of conflicting requirements. To reach that goal, a few of subgoals are set for this dissertation.

1. To verify that the scheme based on SSA can keep the advantages of the SVD-based technique.

2. After we verify that the SSA-based scheme also has the good properties as those of the SVD-based scheme, we have to show that the proposed SSA-based scheme has parameters which can be adjusted to obtain a better performance.

3. To find a relation between those parameters and their physical meanings, and to make a connection between them and conventional analyses. Because all psychoacoustic models analyze signals by using the conventional analysis techniques such as the Fourier transform.

4. To verify that the proposed schemes can deploy information from the perceptual model to improve its performance.

5. To show that the proposed SSA-based scheme can be applied to various applications.

In addition, we also aim to solve a so-called *frame synchronization problem*, the problem that the extraction process requires to know frame positions in advance. All SVD-based schemes are frame-based, and most of them ignore this problem, whereas a few of them adopt synchronization codes. In this work, we exploit the characteristics of the singular spectrum to detect watermarked frames automatically.

1.3 Thesis outline

The organization of this dissertation is shown in Fig. 1.1. The rest of this dissertation consists of five chapters and is organized as follows.

Chapter 2 introduces background knowledge about audio information hiding, its general applications, and some conventional and famous techniques. Since the proposed framework is closely related to methods based on SVD, the SVD-based frameworks are reviewed and analyzed carefully and thoughtfully in order to recognize their advantages and disadvantages. The basic tools and principles necessary to implement the proposed schemes and to investigate their properties, such as singular-spectrum analysis, differential evolution, and psychoacoustic principles, are also provided in this chapter.

Chapter 3 describes the proposed audio-information-hiding frameworks based on SSA. We start with the simplest one, i.e., the core structure, on which the other improved schemes are based. Thenceforth, the more complex schemes are introduced. We show in this chapter how to make a balance between inaudibility and robustness by adopting the differential evolution and how to achieve good performance in transparency by integrating a psychoacoustic model to the scheme. Issues about embedding locations, the effect of embedding the watermark into high- and low-order singular values, and the concept of embedding repetitions, are discussed. In addition, we also propose a novel automatic-frame-detection method without embedding additional synchronization code.

Chapter 4 reports results from implementations and evaluations three related, but different, models based on the frameworks described in the previous chapter. We start

with a database, conditions, and evaluation methods in the first section, after that the parameters that are used to implement each model are given in detail. The automatic frame detection and the audio-information-hiding scheme based on SSA in discrete-wavelet-transform domain are also implemented and investigated their properties in some aspects. Finally, we demonstrate the scheme in which a psychoacoustic model is incorporated.

Chapter 5 gives three examples of applications of the SSA-based audio information hiding: the ownership protection, the information carrier, and the fragile audio-watermarking. Each application has its own requirements, and some of their requirements are different. For example, the fragile audio-watermarking requires fragility, but the ownership protection requires robustness. On the contrary, the robustness is not a concern for the information carrier. These applications are a good example that shows the flexibility of the proposed framework. Evaluations with respect to each application's requirements are performed, and results are reported.

Chapter 6 summarizes this work and emphasizes its contributions to this research field as well as to other research fields. Since the ultimate goal of audio information hiding has yet to achieve, it discusses room for improvement.

1.4 Summary

The unique, innovative points can be summed up as follows: (1) this research exploits the strength of audio watermarking based on SVD but overcomes its drawbacks by proposing a framework based on SSA, (2) using SSA to interpret singular values and to hide information has never been proposed before, (3) the psychoacoustic principles are integrated to the SSA-based framework, (4) the links between SSA and standard analyses need to be established, and (5) the characteristics of the singular spectrum can be used to automatically detect watermarked frames.

In short, this chapter began with giving specific answers to the following questions: what the problem that we want to solve is, why it is worth solving, and whether it is challenging. Then, the motivation and goal of this dissertation are clarified. Lastly, the structure of this thesis is outlined.

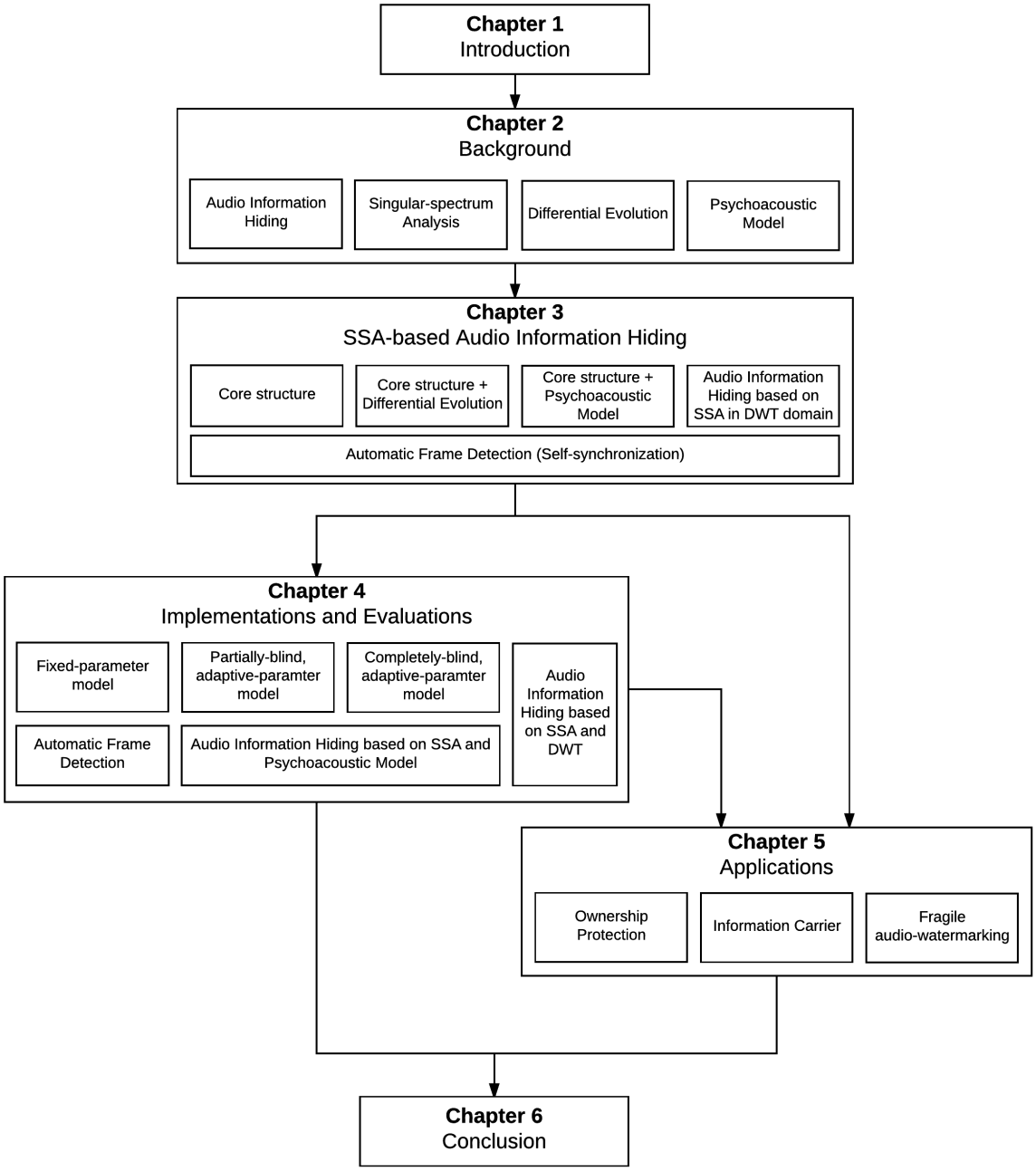

Figure 1.1: Organization of this thesis.

Chapter 2

Background

This chapter introduces background knowledge about audio information hiding, its general applications, and some conventional and famous techniques. Since the proposed framework is closely related to the methods based on SVD, the SVD-based frameworks are reviewed and analyzed carefully and thoughtfully in order to recognize their advantages and disadvantages. The basic tools and principles necessary to implement the proposed schemes and to investigate their properties, such as singular-spectrum analysis, differential evolution, and psychoacoustic principles, are also provided in this chapter.

2.1 Audio information hiding: state of the art

The state of the art of Audio Information Hiding (AIH) is provided in this section. We define AIH and answer the following questions: what it is for, what it has been done so far, and what the clues that we can use to tackle the problem are.

2.1.1 Overview of AIH systems

Strictly speaking, the words information hiding, watermarking, and steganography are different but closely related. Both watermarking and steganography are information hiding but with different aspects. Steganography is the art of hiding information, with a requirement that the existence of the hidden information must be secret [44]. Whereas, this requirement is not the main concern for watermarking. Instead, the relation between the host (the thing into which the hidden information is embedded) and the hidden in-

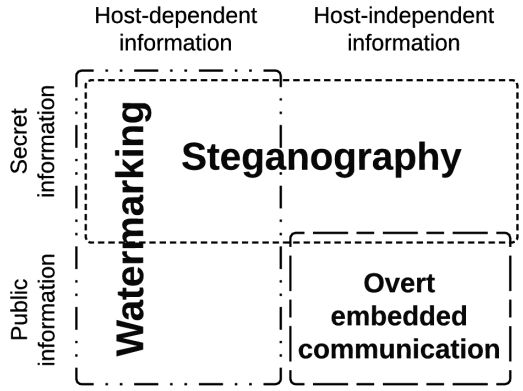

Figure 2.1: Category of information hiding.

formation is more concerned in the watermarking. That is, if the hidden information depends upon the host, then the art of hiding such information into the host is called watermarking [8]. Figure 2.1 depicts the information hiding class. It can be clearly seen that there are both overlapped and separated areas between the steganography and the watermarking.

In most published technical papers, the content and secrecy of the hidden information are not assumed, i.e., they could be any. Thus, the terminologies are used interchangeably. In this work, those terms are interchangeable as well. Accordingly, we also use the words hidden information and watermark as synonym of each other.

Basically, an AIH system consists of two main processes: embedding and extraction, as shown in Fig. 2.2. The embedding process can be considered as a function that takes two inputs, which are a host signal and the hidden information, and returns a watermarked signal. Given the host signal A and the hidden information w, the watermarked signal A^* can be expressed mathematically by the equation $A^* = A + f(A, w)$ [45]. The function f is an embedding function. The extraction process extracts the hidden information $\hat{w}$ from the watermarked signal A^*. The process can be expressed mathematically by the equation $\hat{w} = g(A^*, c(A))$, where the function g is an extracting function and $c(A)$ is a function representing some information that depends on the host signal A. If there is no such information ($c(A) = 0$), the extraction process is called the blind detection; otherwise, the non-blind detection.

In addition to the standard view (Fig. 2.2), the system can also be seen as a com-

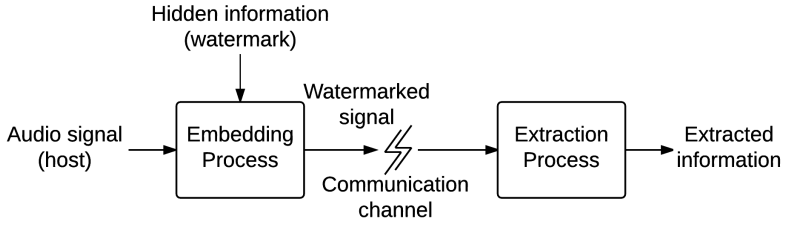

Figure 2.2: Audio-information-hiding system.

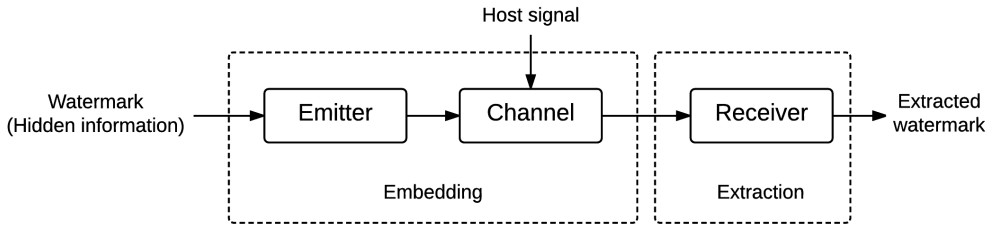

Figure 2.3: Audio information hiding viewed as a communication problem.

munication problem, as shown in Fig. 2.3 [1]. The hidden information is transmitted via the noisy channel. The host audio signal is considered as a noise of the communication channel. The objective is to send the information without distortion, with one additional condition, i.e., the information itself should not be perceived by the human auditory system. This view is useful in the development of digital audio watermarking methods based on the spread spectrum technique [1, 14, 46].

AIH systems can be characterized by a number of properties, such as inaudibility, data payload, blind or non-blind detection, false positive rate, robustness, security, cipher keys, and multiple watermarks [8]. This research focuses on three key properties: *inaudibility*, *robustness*, and *blind detection*. However, the data payload, security, and the multiple watermarks are investigated as secondary properties as well.

2.1.2 Applications of AIH systems

The use of AIH systems for the purpose of copyright control was the original goal of hidden information [16]. Soon after the first successful implementation of a spread-spectrum-based watermarking technique [47], a number of potential applications have been proposed [8, 13, 14, 48, 49]. In this section, we examine some proposed and actual applications.

1. *Ownership protection* or *owner identification.* In the case of dispute of ownership of signals, the owner who knows the watermark hidden inside the signals in question can show the existence of this watermark. Then, he/she can claim the signals are his/her. This application requires a very small false positive detection, the secrecy of the watermark, and the robustness.

2. *Proof of ownership.* This application is related to the previous one, but more demanding. In this scenario, we assume that an attacker can detect and edit the host signal so that the attacker can embed his/her own watermark into the watermarked signal. Thus, the task is not just to identify but to prove the ownership.

3. *Integrity verification* or *content authentication.* The aim of this application is to check whether a signal has been edited since the hidden information was embedded into it. In this scenario, the watermark should be fragile to signal processing, so that the complete extracted watermark can be used to verify integrity [50, 51]. The required properties are the blind detection, the data capacity (normally, the capacity is higher than that required in the previous applications), and the fragility.

4. *Broadcast monitoring.* The monitoring system can be categorized into two groups: passive and active. For the passive monitoring, all received signals are compared with a database, so it raises many problems, such as searching, storing, and managing database [8]. To avoid these problems, the active monitoring system decodes the identification information, which is transmitted along with the broadcast content. Instead of inserting the information in a separate area of the broadcast signal, such as in the vertical blanking interval (VBI) of a video signal, AIH techniques suggest to embed it into the broadcast signal [52–54]. This application requires the high data capacity, the blind detection, and the low rate of false positive.

5. *Fingerprinting* or *transaction tracking.* In the scenario that each distributed signal contains a unique watermark, the watermark can be used to trace the responsible person for misuses or illegal distribution of signals [55, 56]. The requirements for this application are the same as those for the proof-of-ownership application.

6. *Copy control and access control.* A number of methods for copy control, which is integrated in recording or playback systems, have been proposed [15, 57–59]. In

these methods, the embedded watermark serves a useful purpose as a copy control or access control policy. This application requires the blind detection and the low rate of false positive.

7. *Information carrier* or *added value services*. In this scenario, the watermark is not related to copyright information. It adds a value to the host signals, such as annotation and linking content to the Internet [60]. Therefore, malicious attacks are not an issue. The legacy enhancement [8] can be considered as one example of the information carrier. The requirements for this application are the blind detection and the high data capacity.

8. *Applications of steganography*. There are a lot of motivations for two parties who want to communicate secretly [61, 62]. Thus, there are proposed steganography applications, such as the steganography for dissidents or criminals [8]. The common properties for these applications are the high data capacity, the robustness, and the secrecy of the hidden information.

2.1.3 AIH techniques

There are many ways to classify AIH techniques since there are a number of properties that we can use to characterize the algorithm. In other words, classifications depend upon a set of criteria. For example, the audio watermarking techniques can be classified into four categories [63]. The first category embeds information in the time domain, such as the least-significant-bit (LSB) replacement-based schemes. The second category embeds the watermark by introducing an echo to the host signal. The third category embeds information in a certain transform domain, and, the last one, the watermark is embedded based on an audio content and the human auditory system. However, this classification is somehow arbitrary because there are a number of methods that fall into more than one category. For example, the method based on the low-frequency amplitude modification [64] embeds information in the time domain and is based on the human auditory system, so it falls into both the first and the last categories. There is a method based on the FFT amplitude interpolation and the human auditory system [20], thus it falls into the third and the fourth categories. To avoid such a confusion, in this section,

we try not to group the existing AIH techniques. If necessary, a dichotomy is preferred. For example, one can classify the AIH techniques into two groups: the one that deploys properties of the human auditory system [17, 65, 66] and the one that does not [67, 68].

The following subsections briefly explain some conventional and famous AIH techniques.

Least significant bit coding

The least-significant-bit (LSB) coding is one of the earliest and the simplest techniques [69–71]. Basically, the watermark is embedded by the alternation of certain bits of audio samples. The replaced bits are usually the LSBs so as to guarantee inaudibility. Since the most of the information in a sample is contained in the most significant bits, those LSBs hardly affect the sound perception of humans [72]. The extraction process extracts the watermark by decoding the values of those bits.

The LSB-based technique has two major advantages. First, it has a very small computational complexity [13]. So, it is good for real-time applications. Second, it has a very large embedding capacity. The maximum possible capacity is the same as the sampling rate, where all samples are used to embed one bit. However, it has a very serious problem in terms of robustness because the random changes of LSBs, such as adding the white Gaussian noise, destroy the watermark.

Echo hiding

The echo hiding technique was firstly described by Bender et al. in 1996 [73]. The fundamental principle is that humans cannot perceive echoes with a sufficiently short time [74]. Thus, hidden information can be embedded into the host by adding echoes. The watermark bit is encoded by using the difference in delay times and amplitudes of the echoes. The embedding process can be seen as a system that has two possible functions, called *kernels*, as shown in Fig. 2.4(a). The watermark is extracted by detection of spacing between echoes by examining the magnitude at two locations of the autocorrelation of the cepstrum of watermarked signals [73].

From then on, there have been a lot of proposed improved kernels. For example, the dual echo kernel shown in Fig. 2.4(b) was proposed to enhance the detection rate, and

the successive echoes also make the sound quality of echoes signal improved considerably [75]. The backward and forward kernel shown in Fig. 2.4(c), in some sense, is non-causal because the forward echo is added to the host signal before the host signal exists [76]. However, its evaluation results show a lot of advantages, such as the reduction in the echo strength and the increase in detectability. The combined kernels shown in Fig. 2.4(d) combine the dual and backward-and-forward kernels to achieve the imperceptibility and robustness, especially against an echo addition attack [77]. Besides proposing a new kernel, there are other directions of development, such as the adaptive echo hiding proposed to determine the maximum decay rates of impulses in the echo kernels with respect to energy of segments [78].

The echo hiding is blind and robust, but the hidden information can be easily detected by attackers. Moreover, adding echoes has a number of constraints in inaudibility due to the sensitivity of the human auditory system.

So far in this section, we intentionally avoid mentioning a group of AIH techniques, which is based solely on mathematical manipulation of the algebraic features called the *singular values*. That is the techniques based on SVD. As stated in the first chapter, our proposed SSA-based AIH scheme is closely related to the SVD-based ones. In order to keep the advantages of SVD-based methods and to overcome their drawbacks, the SVD-based techniques are reviewed carefully and thoughtfully. The summarized results are given in detail in the following section.

2.1.4 SVD-based audio watermarking

The SVD-based audio watermarking was first reported in 2005 by Özer et al. [25] after the success of applying the same technique in the image watermarking a few years before [79–81]. Since then, it has been a hot research topic in the state-of-the-art audio watermarking techniques [19]. All SVD-based watermarking techniques embed the hidden information into the host signals by means of singular-value modification. The singular value produced by the SVD has interesting properties due to its robust nature [24], such as the sound quality of the audio signal is not affected much by changing singular values to some extent, and the singular values are somewhat stable after various types of signal processing attacks

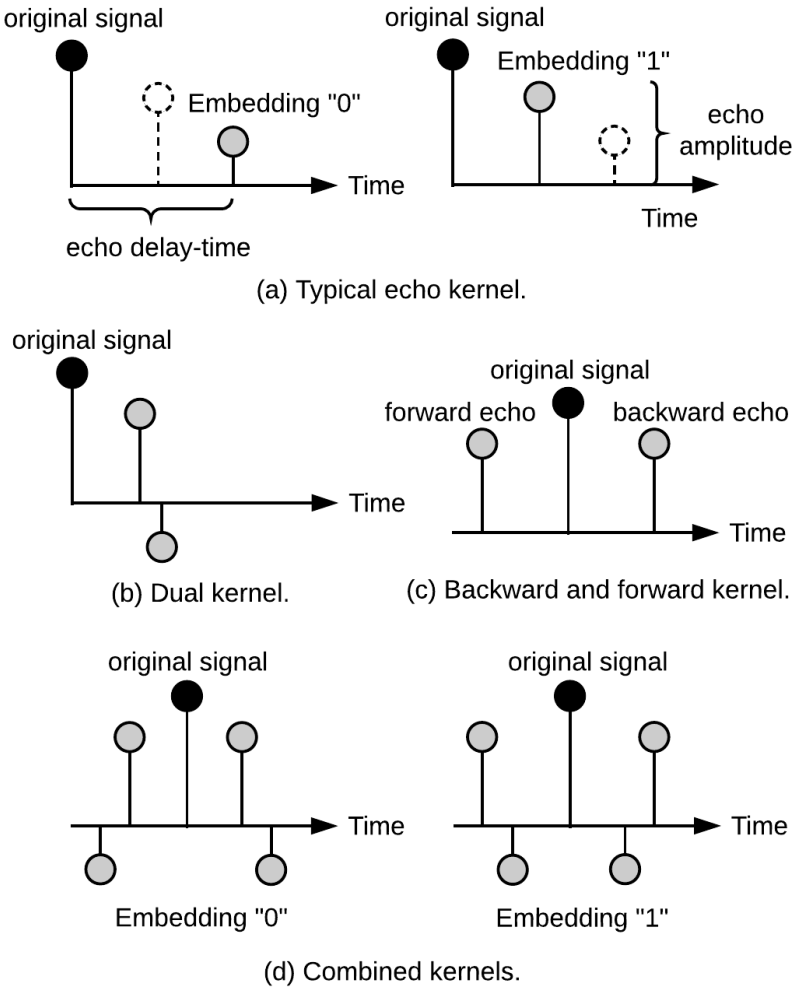

Figure 2.4: Examples of echo kernels.

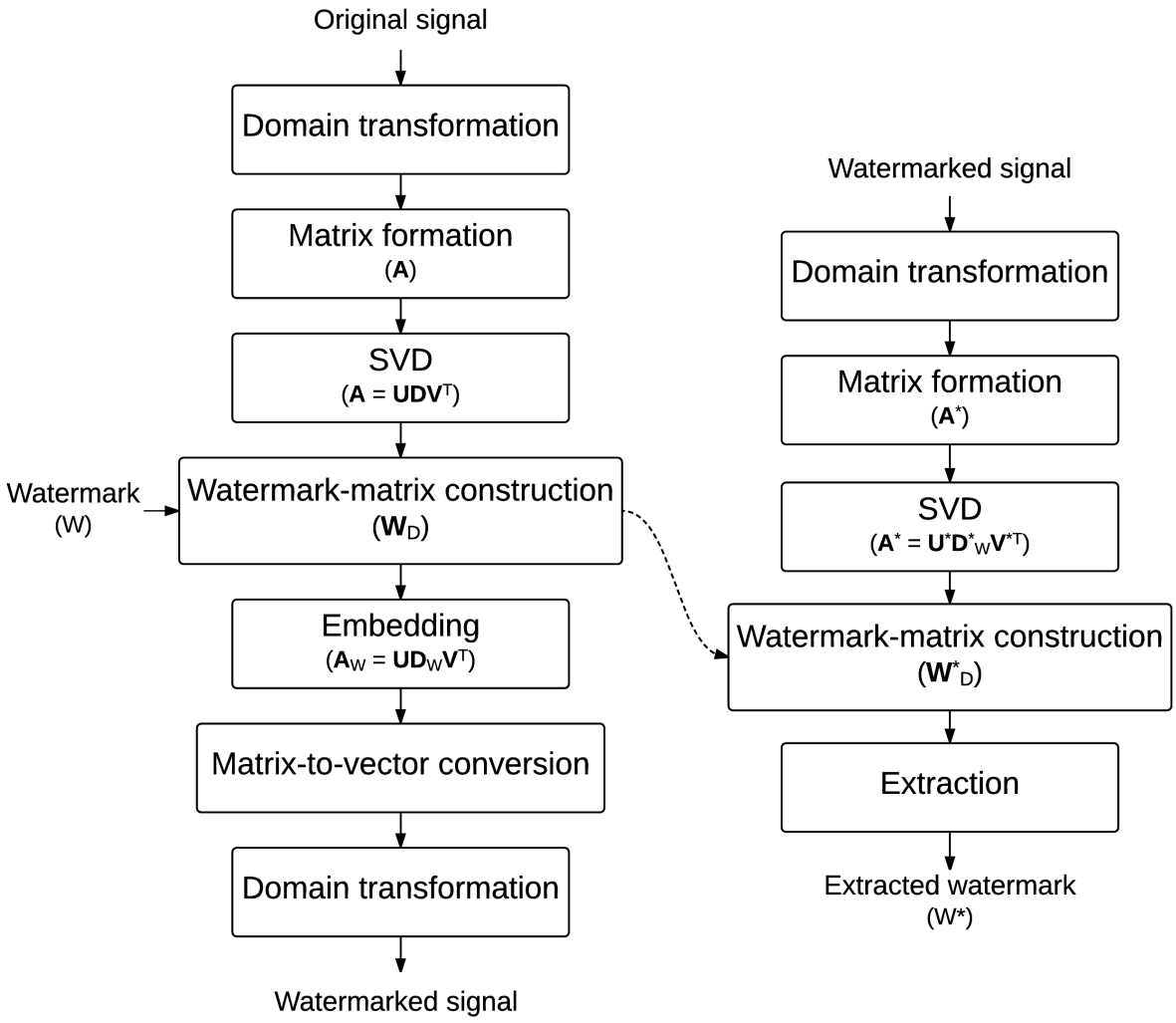

Figure 2.5: Embedding and extraction processes of the SVD-based audio watermarking of the Framework 1.

[33–36]. Thus, the reported experimental results from the SVD-based audio watermarking were promising and impressive in terms of robustness [18, 23, 25–43].

Based on our survey and analysis, the published SVD-based audio watermarking methods can be categorized into two frameworks: *framework 1* [24,25,28,29,37] and *framework 2* [18, 19, 23, 26, 27, 30–36, 38–43]. The embedding and extraction processes of these two frameworks are shown in Figs. 2.5 and 2.6, respectively. The key difference between them is the use of the side information from the embedding process in the extraction process. In order to extract the watermark, the extraction process of *the framework* 1 uses some information which is an output from the watermark-matrix construction of the embedding

process, whereas that of *the framework* 2 does not use such information. Therefore, the methods based on the framework 1 are certainly non-blind. However, some of the methods based on the framework 2 are blind [19,23,26,27,38,82], and some are non-blind [18,30–36].

Framework 1

The left flowchart of Fig. 2.5 illustrates the embedding process. The audio signal represented in a certain domain, such as time or frequency domain, is mapped to the matrix $\boldsymbol{A}$. Then, the matrix $\boldsymbol{A}$ is decomposed by SVD: $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}$, where $\boldsymbol{D}$ is a diagonal matrix, and its diagonal members $\sqrt{\lambda_i}$, called the singular values, are sorted in descending order. The matrix $\boldsymbol{D}$ and the watermark W are used to construct the watermark matrix $\boldsymbol{W}_D$, i.e., $\boldsymbol{W}_D=f_1(\boldsymbol{D},W)$. For example, $\boldsymbol{W}_D=k\times W+\boldsymbol{D}$, where W is a binary image which is encrypted by a chaotic algorithm and k is a scale factor [28]. To determine the scale factor k, a search algorithm, such as the adaptive tabu search (ATS), might be adopted [29].

Then, SVD is performed to the watermark matrix $\boldsymbol{W}_D$: $\boldsymbol{W}_D=\boldsymbol{U}_W\boldsymbol{D}_W\boldsymbol{V}_W^{\mathrm{T}}$. To embed the watermark, the matrix $\boldsymbol{D}$ is replaced by the matrix $\boldsymbol{W}_D$. Thus, the watermarked matrix $\boldsymbol{A}_W$ is the product of the matrices $\boldsymbol{U}$, $\boldsymbol{W}_D$, and the transpose of the matrix $\boldsymbol{V}$. Finally, the watermarked matrix $\boldsymbol{A}_W$ is mapped to a one-dimensional signal, and then it is transformed to the signal in the time domain.

The extraction process is illustrated in Fig. 2.5 (right). The watermarked signal is firstly represented in the certain domain and mapped to the matrix $\boldsymbol{A}^*$. Then, SVD is performed to the matrix $\boldsymbol{A}^*$: $\boldsymbol{A}^*=\boldsymbol{U}^*\boldsymbol{D}_W^*\boldsymbol{V}^{*\mathrm{T}}$. The diagonal matrix $\boldsymbol{D}_W^*$ along with the matrices $\boldsymbol{U}_W$ and $\boldsymbol{V}_W$ obtained from the embedding process are used to construct the watermark matrix $\boldsymbol{W}_D^*$, where $\boldsymbol{W}_D^*=\boldsymbol{U}_W\boldsymbol{D}_W^*\boldsymbol{V}_W^{\mathrm{T}}$. Finally, the watermark matrix $\boldsymbol{W}_D^*$ and the matrix $\boldsymbol{D}$ are used to determine the extracted watermark W^*. Although the host signal does not directly present as an input of the extraction process, the matrix $\boldsymbol{D}$, which contains some information of the host signal, is required. Therefore, the methods based on this framework are considered as non-blind.

Based on this framework, many methods develop differently in the domain transformation, matrix formation, and the watermark-matrix construction. For example, the short-time Fourier transform [25], discrete wavelet transform [40], or discrete cosine transform [28] might be adopted as the domain transformation. The analysis of matrix for-

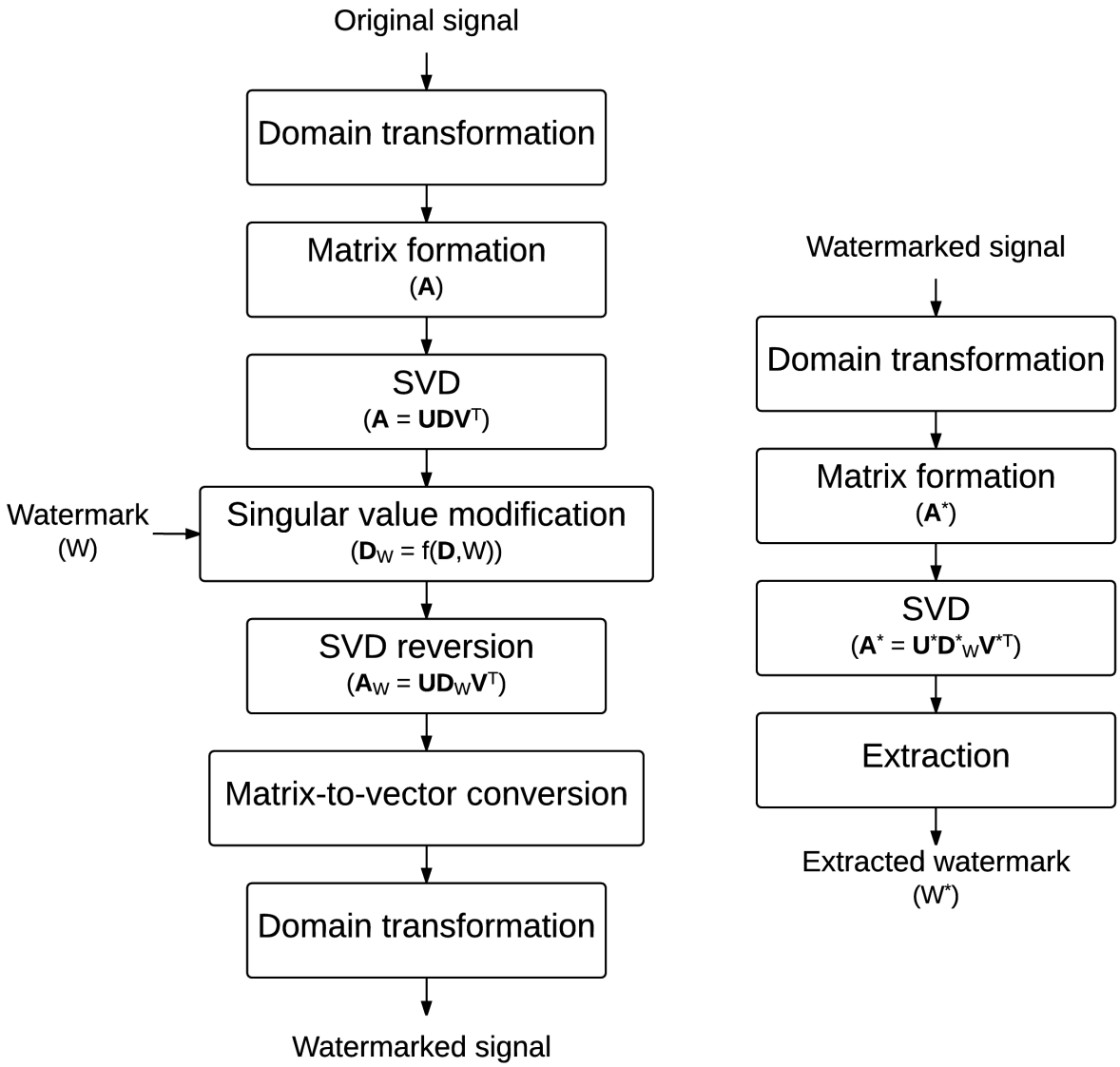

Figure 2.6: Embedding and extraction processes of the SVD-based audio watermarking of the Framework 2.

mation is discussed in [83]. The watermark-matrix construction (the function f_1) in the embedding process directly affects the extraction. For example, if $\boldsymbol{W}_D = k \times W + \boldsymbol{D}$, then the extracted watermark W^* is $\frac{\boldsymbol{W}_D^* - \boldsymbol{D}}{k}$ [28, 29].

It should be noted that, even though the experimental results from the methods based on this framework are excellent in terms of robustness, the high bit-detection rates are possibly due to the false positive detection [24, 84, 85]. The false-positive rate raises because the matrices $\boldsymbol{U}_W$ and $\boldsymbol{V}_W$, which contain most information of the watermark, are used to construct the watermark matrix $\boldsymbol{W}_D^*$, and our preliminary simulations have confirmed this effect.

Framework 2

The embedding process is shown in Fig. 2.6 (left). In contrast with the framework 1, this framework performs SVD only once in the embedding process, i.e, the SVD is performed to the matrix $\boldsymbol{A}$, which represents the original signal in a certain domain: $\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}$. Then, the watermark W is embedded by modification of the diagonal matrix $\boldsymbol{D}$, i.e., $\boldsymbol{D}_W = f_2(\boldsymbol{D}, W)$, where $\boldsymbol{D}_W$ is the modified $\boldsymbol{D}$. The function f_2 is called the singular-value modification rule. The watermarked matrix $\boldsymbol{A}_W$ is constructed by the SVD reversion where the diagonal matrix $\boldsymbol{D}$ is replaced with the matrix $\boldsymbol{D}_W$, i.e., $\boldsymbol{A}_W = \boldsymbol{U}\boldsymbol{D}_W\boldsymbol{V}^{\mathrm{T}}$. Finally. the matrix $\boldsymbol{A}_W$ is conversed and transformed to the time-domain watermarked signal.

The extraction process is shown in the right flowchart of Fig. 2.6 (right). The watermarked signal is transformed and mapped to the matrix $\boldsymbol{A}^*$. Then, the matrix $\boldsymbol{A}^*$ is decomposed by SVD, and the diagonal matrix $\boldsymbol{D}_W^*$ is obtained. The extracted watermark W^* is decoded from the matrix $\boldsymbol{D}_W^*$ with respect to the singular-value modification rule f_2. The extraction process may require some information of the host signal as an input [18, 30–36], or it may not [23, 26, 27, 38, 82]. Thus, the methods based on this framework can be either blind or non-blind. For example, let us consider the modification rule of the methods [18, 30, 31].

Given a watermark bit $w \in \{0, 1\}$, the first singular value $\sqrt{\lambda_1}$ of the matrix $\boldsymbol{A}$ is replaced with modified singular value $\sqrt{\lambda_{1W}} = \sqrt{\lambda_1} \times (1 + \alpha \cdot w)$, where α is a scale factor called the watermark intensity. According to this rule, the extracted watermark bit w^*

is 1 if $\frac{\sqrt{\lambda_{1W}}}{\sqrt{\lambda_1}} = 1+\alpha$, and the w^* is 0 if $\frac{\sqrt{\lambda_{1W}}}{\sqrt{\lambda_1}} = 1$. It can be seen clearly that, in order to extract the watermark bit, these methods require the first singular value $\sqrt{\lambda_1}$ of the matrix $\boldsymbol{A}$, which represents the host signal. Note that different methods may modify other singular values. For example, some methods [18, 30, 31, 33, 35, 36] modify only the largest singular value. Other methods [23, 26, 27, 32, 38] modify all singular values, and there is a method [82] that modifies only some small singular values.

The quantization index modulation (QIM) can be applied to the singular-value modification to make the methods to be blind [23, 26, 27].

Remarks

1. To avoid the false-positive-detection problem as happened to the *framework* 1, any information of the watermark should not be used in the extraction process.
2. All SVD-based audio watermarking methods have shown that the singular values are stable to some extent because the watermark encoded in the singular values is extractable after several signal processing attacks, as evidenced by the published results. Thus, the singular value is proved to be a potential candidate for hiding information. However, to the best of our knowledge, all SVD-based audio watermarking methods have treated an audio signal as a meaningless time-series, and they have yet to provide the insight or explanation of its effectiveness. The question of interpretation of the singular value has never been discussed and answered. Moreover, all modification rules are uninformed, i.e., the nature of signals or the nature of signal perception have never been taken into account. The sole philosophy of all SVD-based methods so far seems to be that, if something changes very slightly, we hardly notice.

2.2 Singular-spectrum analysis

Singular-spectrum analysis (SSA) is a time-series analysis technique, which is useful for identifying and extracting oscillatory components from a signal [86]. Compared to other techniques for investigating time-series data, such as performing a Fourier transform or the principle component analysis, the SSA is much younger. It was proposed by Broomhead and King in 1986 [87] and has been widely used in various applications, such as

extraction of periodicities and finding structures in time series [88]. As evidenced by becoming a standard tool in the analysis of climate, meteorological, and geophysical data, it has proved to be one of the successful techniques [86]. It is also popular for analyzing biomedical signals [89–91].

The SSA is a model-free algorithm in the sense that, during analyzing, statistical assumptions concerning the signal are not made, so that it can be applied to arbitrary signals including non-stationary ones [86]. There are many types of SSAs. The following subsection describes the basic SSA, on which our proposed scheme is based.

2.2.1 The basic SSA

The basic SSA consists of four steps: embedding, singular value decomposition (SVD), grouping, and diagonal averaging. The first two steps form the decomposition (or analysis) stage. The last two steps form the reconstruction (or synthesis) stage.

Embedding step

A signal $F = [f_0\ f_1\ f_2\ ...\ f_{N-1}]^{\mathrm{T}}$ of length N, where $N > 2$, is mapped to the trajectory matrix $\boldsymbol{X}$ of size $L \times K$.

$$\boldsymbol{X} = \begin{bmatrix} f_0 & f_1 & f_2 & \cdots & f_{K-1} \\ f_1 & f_2 & f_3 & \cdots & f_K \\ f_2 & f_3 & f_4 & \cdots & f_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_{L-1} & f_L & f_{L+1} & \cdots & f_{N-1} \end{bmatrix}, \tag{2.1}$$

where $K = N - L + 1$, and where L is the only parameter of the basic SSA, called a *window length* of matrix formation, and has a maximum value of N. For $j = 0, 1, ..., K-1$, let $\boldsymbol{x}_j$, called a lagged vector, denote the j^{th} column of the matrix $\boldsymbol{X}$, i.e., $\boldsymbol{x}_j = [f_j\ f_{j+1}\ ...\ f_{j+L-1}]^{\mathrm{T}}$. Thus, $\boldsymbol{X} = [\boldsymbol{x}_0\ \boldsymbol{x}_1\ \boldsymbol{x}_2\ ...\ \boldsymbol{x}_{K-1}]$. Since the trajectory matrix $\boldsymbol{X}$ has equal elements on the minor diagonals (ascending skew-diagonals from left to right), it is a Hankel matrix.

SVD step

The trajectory matrix $\boldsymbol{X}$ is decomposed into a product of three matrices $\boldsymbol{U}$, $\boldsymbol{D}$, and $\boldsymbol{V}$ with the following relationship.

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{V}^{\mathrm{T}}, \tag{2.2}$$

where the columns of $\boldsymbol{U}$ and $\boldsymbol{V}$, $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ for $i = 0, 1, ..., L-1$, are eigenvectors of $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$ and $\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}$, respectively, and $\boldsymbol{u}_i$ and $\boldsymbol{v}_i$ are sorted in the descending order of the corresponding eigenvalues of $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$, and where $\boldsymbol{D}$ is the diagonal matrix whose elements are the square root of the eigenvalues, called the singular value.

Let $\lambda_0, \lambda_1, ...,$ and λ_{L-1} denote the eigenvalues of $\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}$. The trajectory matrix $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2 + ... + \boldsymbol{X}_d, \tag{2.3}$$

where $\boldsymbol{X}_k = \sqrt{\lambda_k}\boldsymbol{u}_k\boldsymbol{v}_k^{\mathrm{T}}$ and $d = \arg\max_k (\lambda_k > 0)$.

Grouping step

The set of indices of $\boldsymbol{X}_k$, $\{1, 2, ..., d\}$, obtained from the previous step is partitioned into m disjoint subsets I_l for $l = 1, 2, ..., m$. Then, $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, ..., and $\boldsymbol{X}_d$ are grouped into m groups, so that the trajectory matrix $\boldsymbol{X}$ can be written as

$$\boldsymbol{X} = \boldsymbol{X}_{I_1} + \boldsymbol{X}_{I_2} + ... + \boldsymbol{X}_{I_m}. \tag{2.4}$$

Note that, ideally, all groups $\boldsymbol{X}_{I_l}$ should be Hankel matrices or, at least, satisfy the separability conditions [86].

Diagonal averaging step

Each matrix $\boldsymbol{X}_{I_l}$ is mapped by diagonal averaging (Hankelizing) to a new time-series of length N. The Hankelization of a matrix $\boldsymbol{Y}$ of dimension $L \times K$ to the time-series $G = [g_0\ g_1\ g_2\ ...\ g_{N-1}]^{\mathrm{T}}$ is defined as follows.

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y^*_{m,k-m+2}, & \text{for} \quad 0 \le k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y^*_{m,k-m+2}, & \text{for} \quad L^* - 1 \le k < K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y^*_{m,k-m+2}, & \text{for} \quad K^* \le k < N, \end{cases} \tag{2.5}$$

where y_{ij} is an element at the row i and column j of the matrix $\boldsymbol{Y}$, $L^* = \min(L, K)$, $K^* = \max(L, K)$, $y^*_{ij} = y_{ij}$ if $L < K$, and where $y^*_{ij} = y_{ji}$ if $L \geq K$.

2.2.2 Interpretation of singular value

An example of the SSA decomposition of a signal is shown in Fig. 2.7. The first panel, labeled *Org* on the vertical axis, is the signal zoomed to observe the first 300 samples. The basic SSA is used to decompose the signal of the size of 2450 with the window length of 500. The first 200 singular values are shown in Fig. 2.8. Note that there are more than 200 singular values that are greater than zero, and the number of the positive singular values can be up to the size of the window length L.

The second to sixth panels show the 1^{st}, 5^{th}, 50^{th}, 100^{th}, and the 200^{th} oscillatory components (X_i) of the signal, respectively. These components associate with the 1^{st}, 5^{th}, 50^{th}, 100^{th}, and the 200^{th} singular values, respectively, i.e., X_i is the result of Hankelization of the matrix $\boldsymbol{X}_i = \sqrt{\lambda_i}\boldsymbol{u}_i\boldsymbol{v}_i^{\mathrm{T}}$. Thus, in the light of SSA, the singular value can be interpreted as a scale factor of the oscillatory component. Since singular values are sorted in descending order, the lower the component order, the more contribution to the signal. The result of adding the first 100 oscillatory components is shown at the top panel of Fig. 2.9, and the difference between the original and reconstructed signals is shown in the bottom panel of the same figure. When all components are added up, the original signal is obtained due to the linearity.

2.3 Differential evolution

Differential evolution is a direct search method that solves an optimization problem by iteratively improving candidate solutions with regard to an objective function [92]. The phrase "direct search" means that the method does not employ techniques of classical analysis, i.e., it does not require the objective function to be differentiable. Thus, it is often described as *derivative-free* [93]. Similar to the evolution strategies and genetic algorithms [94], the differential evolution is inspired by biological mechanisms of evolution and is population-based, i.e., it maintains multiple assignments or starting points forming a population [95]. As a result, they can outperform a starting point problem, which occurs

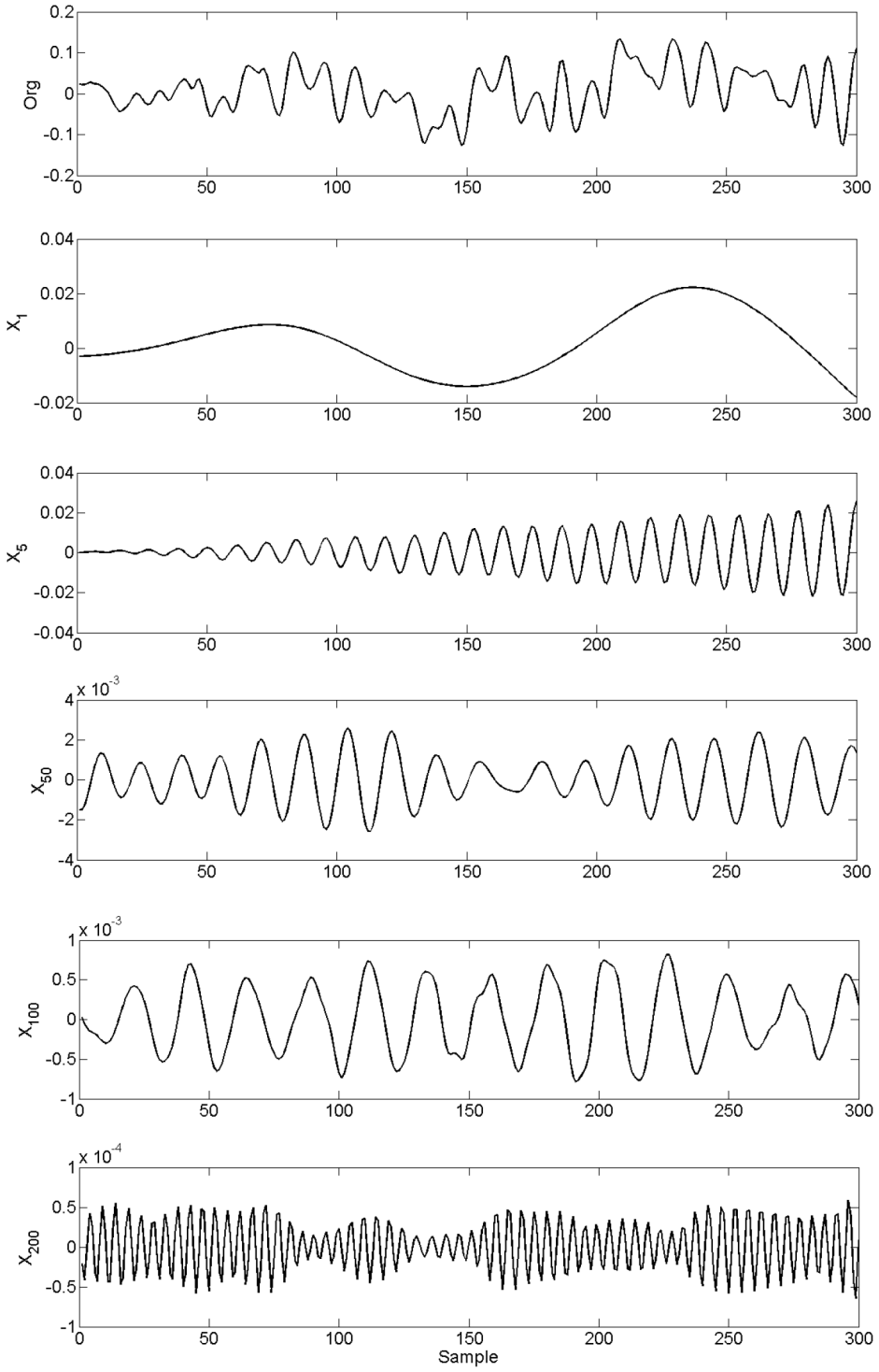

Figure 2.7: Example of using SSA to decompose a signal (top panel) into additive oscillatory components (last five panels).

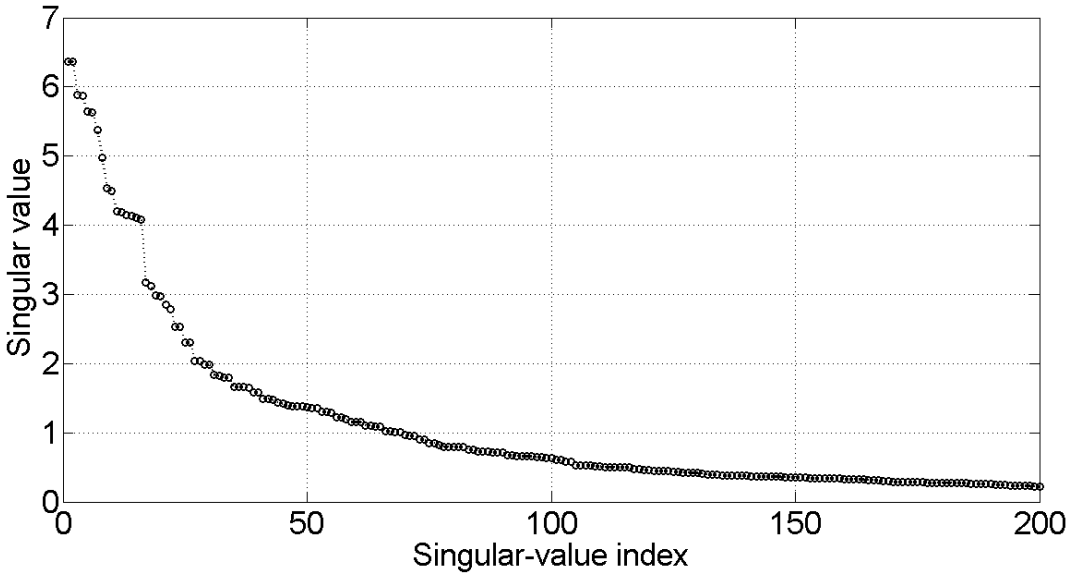

Figure 2.8: Example of the first 200 singular values.

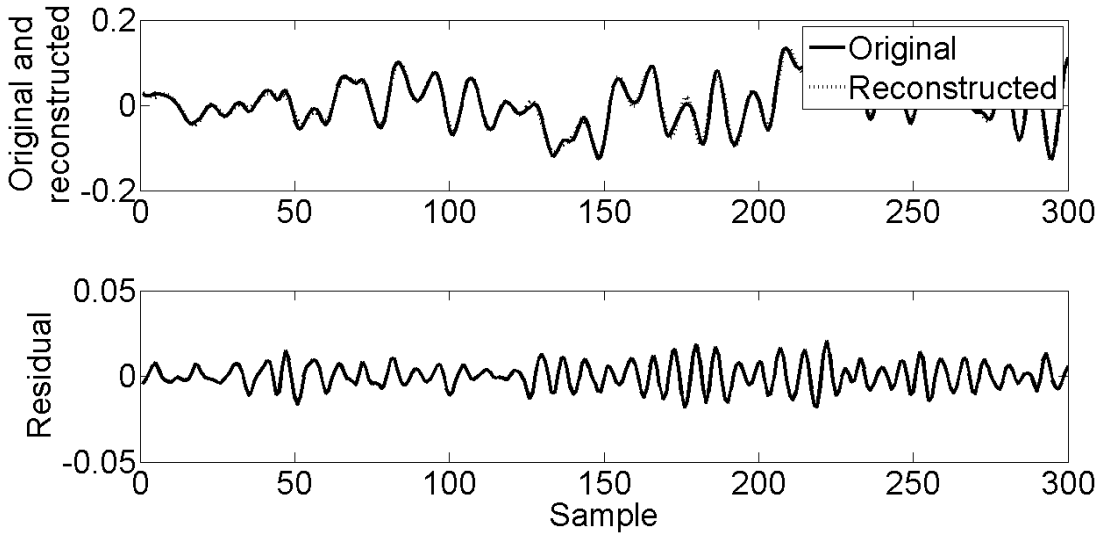

Figure 2.9: Original and reconstructed signals (top). The difference between original and reconstructed signals (bottom).

when an optimizer is stuck in local extrema because of its randomized, initial point.

Collectively, the evolution strategies, genetic algorithms, differential evolution and the likes are classified as *evolutionary algorithms*. However, they are different in many ways [96]. For example, simplest implementations of the evolution strategies still face the step size problem, which refers to the difficulty of determining the suitable step size for searching. The optimal point may be missed if the step size is too large. But if it is too small, computing time increases exponentially due to the curse of dimentionality [97]. The differential evolution gets around this problem by deploying difference vectors. The advantage of adopting difference vectors is that the step size and the orientation adapt to the objective function landscape automatically [96]. Even though more complex implementations of the evolution strategies such as the Nelder-Mead polyhedron search [98] and the controlled random search [99] deploy difference vectors as well, the population size of the Nelder-Mead algorithm is very limited, and the controlled random algorithm exerts too high selective pressure due to its continually-replace-the-worst strategy [96].

Compared to the genetic algorithms, the differential evolution represents parameters differently, i.e., it encodes all parameters as a floating-point number, whereas the genetic algorithms encode them as a bit string. That means arithmetic and logical operators are required in order to manipulate parameter vectors for the differential evolution and the genetic algorithms respectively. In terms of complexity and flexibility, the arithmetic operation is better than the logical one [96]. Moreover, the genetic algorithms typically select base vectors (or parents) with a bias toward better vectors, while the differential evolution does equally. In other words, the genetic algorithms more exert selective pressure than the differential evolution does. It can cause a premature convergence to a local extrema [96].

The differential evolution consists of four processes: initialization, mutation, crossover, and selection, as illustrated in Fig. 2.10. There are many variants of the differential evolution, depending upon the way the base vectors are specified, the number of difference vectors used, and the crossover scheme. The following subsections describe the summary of the classic differential evolution (technically, *DE/rand/1/bin*[1]), which is adopted as a parameter optimizer in the proposed SSA-based audio information hiding.

[1]The technical name for the different variants is the notation *DE/x/y/z*, where x specifies how to choose the base vectors, y is the number of difference vectors used in the mutation process, and z is the

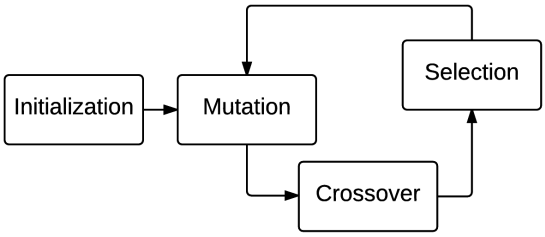

Figure 2.10: Differential evolution processes.

2.3.1 Initialization

Without loss of generality, we assume that the aim of the optimizer is to deliver parameters x_i for $i=0,1,...,D-1$ that minimize a cost function $f(x_0, x_1, ..., x_{D-1})$.

For $j=0,1,...,NP-1$, let $\boldsymbol{x}_{j,G}$, called a target vector, denote D-dimensional parameter vectors forming a population for the generation G, where G is a non-negative integer. A parameter vector $\boldsymbol{x}_{j,G}$ is constructed from the parameters x_i for $i=0,1,...,D-1$, i.e., $\boldsymbol{x}_{j,G}=[x_{0,j,G} \quad x_{1,j,G} \quad ... \quad x_{D-1,j,G}]^{\mathrm{T}}$.

Initialization is the process of initializing and initiating $\boldsymbol{x}_{j,0}$. To initialize, firstly, the maximum $x_{i,\max}$ and the minimum $x_{i,\min}$ for each parameter x_i must be specified. The i^{th} parameter of the j^{th} vector for the first generation, i.e., $x_{i,j,0}$, is defined as follows.

$$x_{i,j,0} = \Omega_i \times (x_{i,j,\max} - x_{i,j,\min}) + x_{i,j,\min}, \tag{2.6}$$

where Ω_i is a random number generator that returns a number from within the interval $[0,1)$. We put the subscript i to indicate that for each parameter x_i a new random number is generated. Note that the random numbers can be uniformly-distributed or Gaussian-distributed.

2.3.2 Mutation

For each target vector $\boldsymbol{x}_{j,G}$, a mutant vector $\boldsymbol{v}_{j,G}$ is created by the following differential mutation.

$$\boldsymbol{v}_{j,G} = \boldsymbol{x}_{r_0,G} + F \times (\boldsymbol{x}_{r_1,G} - \boldsymbol{x}_{r_2,G}), \tag{2.7}$$

crossover scheme. Thus, *DE/rand/1/bin* means that the base vectors are uniformly randomly chosen, only one difference vector is used in the mutation process, and the number of parameters donated to a trial vector by a mutant vector follows a binomial distribution [92, 96].

where r_0 is the based vector index, r_1, and r_2 are called the difference vector indices which are randomly chosen from $\{0, 1, ..., NP-1\}\backslash\{j\}$ and mutually different. $F \in (0, 2]$ is a predefined real number that controls the convergence or the rate the population evolves by scaling a difference between two vectors $\boldsymbol{x}_{r_1,G}$ and $\boldsymbol{x}_{r_2,G}$.

Note that the requirement for four different vector indices: j, r_0, r_1, and r_2, makes the minimum members of population to be four. If some pairs of these parameters are equal, the performance of the differential evolution will drop. For example, if $r_1 = r_2$, then there is no mutation. If $r_0 = j$, then there is only mutation. If r_1 or r_2 is equal to r_0, then the differential evolution is reduced to the arithmetic recombination. Only when they are distinct ($j \neq r_0 \neq r_1 \neq r_2$), the differential evolution achieves the best performance [96].

2.3.3 Crossover

Crossover is the process of making trial vectors $\boldsymbol{u}_{j,G}$ from the target vector $\boldsymbol{x}_{j,G}$ and the mutant vectors $\boldsymbol{v}_{j,G}$ for $j = 0, 1, ..., NP-1$, where some parameters $u_{i,j,G}$ are the copy of $x_{i,j,G}$, and the others are the copy of $v_{i,j,G}$. There are many crossover schemes determining which parameters copied from which vectors, such as the one-point crossover, the N-point crossover, the exponential crossover, and the uniform crossover [96]. Most literature suggested the exponential and the uniform ones [100–103]. The classic differential evolution adopts the uniform crossover, defined as follows.

$$u_{i,j,G} = \begin{cases} v_{i,j,G} & \text{if } (\Omega_i \leq CR) \text{ or } i = i_{\text{rand}} \\ x_{i,j,G} & \text{otherwise,} \end{cases} \tag{2.8}$$

where $CR \in [0, 1]$ is an user-defined crossover constant and can be considered as a mutation rate. Typically, the genetic algorithms set the mutation rate of D^{-1}, but for the differential evolution either $0 \leq CR \leq 0.2$ or $0.9 \leq CR \leq 1$ is recommended [92, 96]. The index i_{rand} is chosen randomly from $\{0, 1, ..., D-1\}$, which ensures that the trial vector $\boldsymbol{u}_{j,G}$ will get at least one parameter from the mutant vector $\boldsymbol{v}_{j,G}$. The random number generator Ω_i is uniform.

2.3.4 Selection

By using a greedy criterion, the trial vector $\boldsymbol{u}_{j,G}$ is compared with the target vector $\boldsymbol{x}_{j,G}$. If the cost function of the trial vector, $f(\boldsymbol{u}_{j,G})$, is not greater than that of the target

vector, $f(\boldsymbol{x}_{j,G})$, the target vector $\boldsymbol{x}_{j,G+1}$ for the next generation $G+1$ is replaced by the trial vector $\boldsymbol{u}_{j,G}$; otherwise, it retains the same vector $\boldsymbol{x}_{j,G}$.

$$\boldsymbol{x}_{j,G+1} = \begin{cases} \boldsymbol{u}_{j,G} & \text{if } f(\boldsymbol{u}_{j,G}) \leq f(\boldsymbol{x}_{j,G}) \\ \boldsymbol{x}_{j,G} & \text{otherwise.} \end{cases} \tag{2.9}$$

Thus, there are NP competitions in a generation. After we have all members $\boldsymbol{x}_{j,G+1}$ for the generation $G+1$, the next mutant vectors $\boldsymbol{v}_{j,G+1}$ are created to be the contributors to the next trial vectors $\boldsymbol{u}_{j,G+1}$ which will be compared with the target vectors $\boldsymbol{x}_{j,G+1}$ for the survival selection for the next generation $G+2$, and so on. The evolutionary cycle of mutation, crossover, and selection iteratively continues until a stopping criterion, such as the maximum number of generations or the predefined lowest cost value, is met. Then, the vector that yields the lowest cost is the solution.

2.4 Human auditory perception and psychoacoustic models

Psychoacoustics is the science of sound perception. It investigates the statistical relationships between acoustic stimuli and hearing sensations and aims to construct the psychoacoustic model [104,105]. We adopt the psychoacoustic model 1, which is used in MPEG-1 Audio [106,107], to our proposed SSA-based scheme. Hence, this section summarizes the important psychoacoustic principles relevant to it.

2.4.1 Absolute threshold of hearing

Absolute threshold of hearing is the minimum detectable sound pressure level in the absent of any other sounds [108]. It is a function of frequency with the following relation.

$$T_q(f) = 3.64\left(\frac{f}{1000}\right)^{-0.8} - 6.5e^{-0.6\left(\frac{f}{1000}-3.3\right)^2} + 0.001\left(\frac{f}{1000}\right)^4, \tag{2.10}$$

where $T_q(f)$ is the threshold of hearing of the sound pressure level (SPL) of a pure tone at frequency f. This relation is plotted and shown in Fig. 2.11. The meaning of this curve is that we cannot hear a pure tone with SPL under the absolute threshold curve [108].

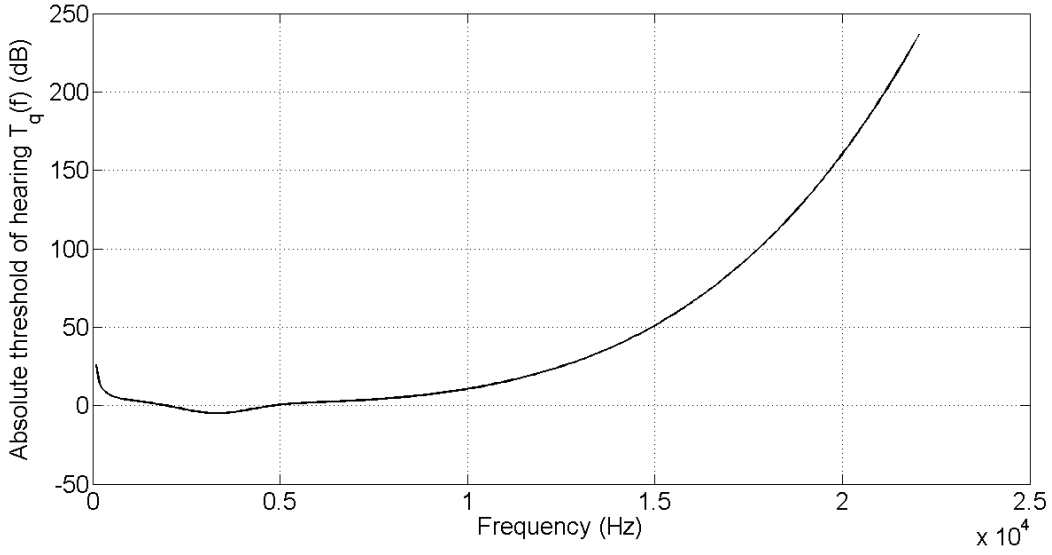

Figure 2.11: Absolute threshold of hearing.

2.4.2 Simultaneous masking

Masking is the situation when the presence of one sound (the masker) renders another sound (the signal or the maskee) less detectable [21]. If the masker and maskee are presented simultaneously, the situation is called simultaneous masking [109]. The difference between the sound pressure levels of the masker and the signal is called the signal-to-mask ratio (SMR).

Since the frequency components of an audio signal can be considered as either noise-like or tone-like [105, 110], both the masker and maskee can be either noise-like or tone-like as well. Therefore, one can classify the masking into four types: tone-masking-tone (TMT), tone-masking-noise (TMN), noise-masking-tone (NMT), and noise-masking-noise (NMN).

The phenomena of TMT and NMN are normally excluded from psychoacoustic models because there are no good models for them [110]. The minimum SMR of TMN (the smallest difference between the sound pressure levels of the pure tone masking the narrow-band noise and the narrow-band noise) is between 21 and 28 dB [111, 112]. In the case of NMT, the minimum SMR is between 2 and 6 dB [109].

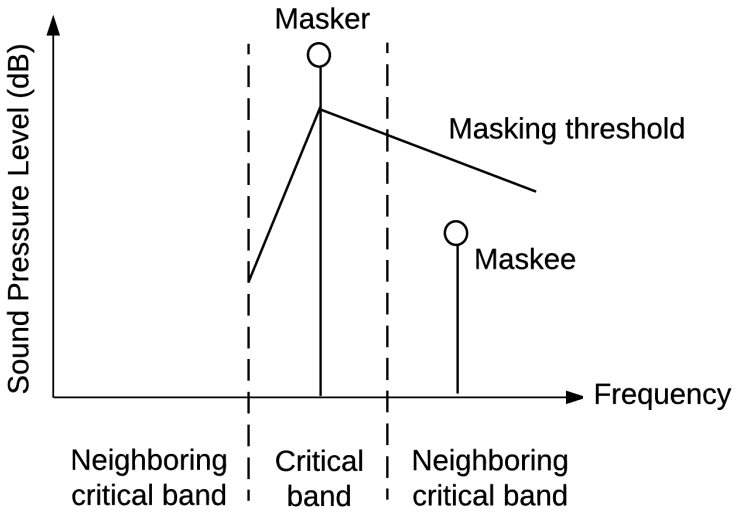

Figure 2.12: Spreading of masking into neighboring critical bands.

2.4.3 Spread of masking

The masking summarized in the previous subsection is the situation when a masker masks maskees within the *critical band*[2]of the masker. However, masking effect can exist when the maskee is in the different critical band. This phenomenon is called the spread of masking. An representation example of the spread of the masking is shown in Fig. 2.12 [113].

2.4.4 Psychoacoustic model

The psychoacoustic model 1 delivers the SMR of the analyzed signal. According to the standard ISO/IEC 11172-3, it consists of 5 steps, as illustrated in Fig.2.13 [105,110]. The overview of these steps is summarized as follows. First, we calculate the fast Fourier

[2]Critical bandwidth is a measure of the effective bandwidth of the auditory filter and is defined empirically by measuring some aspect of perception as a function of the bandwidth of the stimuli and trying to determine a break point in the results [108]. The critical bandwidth $\mathrm{BC}_c(f)$ can be approximated by $25 + 75\left(1 + 1.4\left(\frac{f}{1000}\right)^2\right)^{0.69}$, where f is the center frequency [109].

Note that the critical bandwidth expressed by this equation is widely used in perceptual models for audio coding [105]. However, there is another measure, which is used widely in the current research on psychoacoustics, that estimates the bandwidths of the filters in human hearing, called the equivalent rectangular bandwidth or ERB. In this work, we follow the measure that the original model uses.

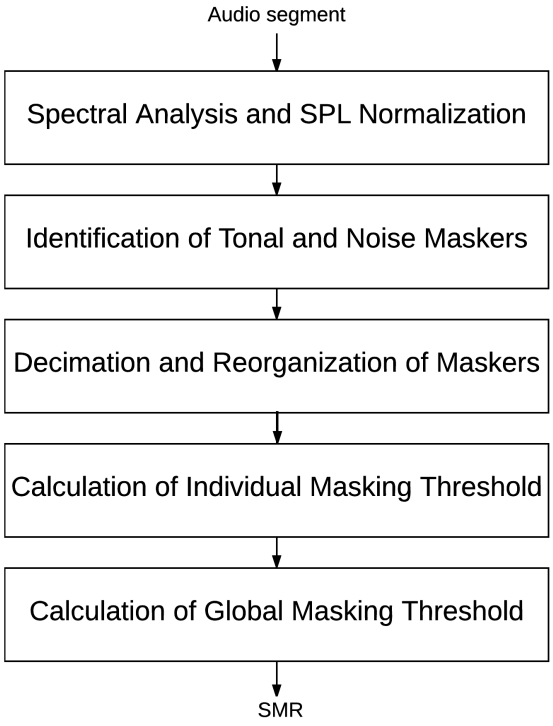

Figure 2.13: Psychoacoustic model 1.

transform (FFT) and the power spectral density (PSD) of the signal. Then, the PSD is normalized to a sound pressure level (PSD) of 96 dB, i.e., the maximum SPL is limited to 96 dB. Second, we use the PSD to identify the tonal and noise components. A component is classified as the tonal one if it is sinusoid-like; otherwise, it is a noise component. After we obtain the tonal and noise components, we calculate the tonal and noise maskers in each critical band. Third, we decimate unnecessary maskers by using two psychoacoustic principles so that we obtain only the relevant ones. Fourth, we use the survival maskers to calculate the individual masking threshold. Finally, we combine all masking thresholds to produce the global masking threshold. The details of these steps are formulated mathematically as follows [105, 110].

1. Input audio sample $s(n)$ is normalized according to the FFT-length N_{FFT} (512 points) and the quantization bit depth b, i.e., the number of bit per sample.

$$x(n) = \frac{s(n)}{N_{\mathrm{FFT}} \times (2^{b-1})}, \tag{2.11}$$

for $n = 0, 1, 2, ..., M-1$, where M is the length of the audio signal. Then, the normalized input $x(n)$ is segmented into frames of 512 samples, and, using a $\frac{1}{16}^{\mathrm{th}}$-

overlapped Hanning windows $w(n)$, the power spectral density $P(k)$ is estimated by using a 512-point FFT.

$$P(k) = 90.302 + 10\log\left|\sum_{n=0}^{N_{\mathrm{FFT}}-1} w(n) \times x(n) \times e^{-j\frac{2\pi kn}{N_{\mathrm{FFT}}}}\right|^2, \tag{2.12}$$

for frequency indices $k=0$ to $\frac{N_{\mathrm{FFT}}}{2}$, and the Hanning window $w(n)$ is defined as

$$w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N_{\mathrm{FFT}}}\right)\right]. \tag{2.13}$$

The output of this step is the normalized PSD $P(k)$.

2. The local maxima in the PSD of input signal obtained from the previous step that are greater than neighboring components within a certain Bark distance by 7 dB are classified as tone-like components. Note that the Bark scale is a frequency scale on which equal distances correspond with perceptually equal distances [114]. The relationship between the Bark scale z and the linear frequency scale f can be expressed by the following formula [109].

$$z = 13\arctan(0.00076 \times f) + 3.5\arctan[(\frac{f}{7500})^2]. \tag{2.14}$$

The tonal component set S_T is defined as follows.

$$S_T = \{P(k)|P(k) > P(k \pm 1) \quad \text{and} \quad P(k) > P(k \pm \Delta_k) + 7\}, \tag{2.15}$$

where

$$\Delta_k \in \begin{cases} [2], & \text{for} \quad 2 < k < 63 \quad (0.17 - 5.5 \text{ kHz}), \\ [2,3], & \text{for} \quad 63 \leq k < 127 \quad (5.5 - 11 \text{ kHz}), \\ [2,6], & \text{for} \quad 127 \leq k \leq 256 \quad (11 - 20 \text{ kHz}), \end{cases} \tag{2.16}$$

Then, the tonal masker $P_{TM}(k)$ of each tonal component is calculated by

$$P_{TM}(k) = 10\log\sum_{r=-1}^{1} 10^{0.1\times P(k+r)}. \tag{2.17}$$

The components not within the $\pm\Delta_k$ neighborhood of the tonal masker are used to calculate the noise masker $P_{NM}(\bar{k})$. Note that there is only one noise masker for each critical band.

$$P_{NM}(\bar{k}) = 10\log\sum_{r} 10^{0.1\times P(r)}, \tag{2.18}$$

for all r that satisfies the relation $P(r) \notin S_T$, where $\bar{k}$ is the frequency index nearest to the geometric mean of each critical band. Given the lower and upper spectral-line boundaries of the critical band, which are v and w, respectively, the geometric mean $\bar{k}$ is defined as

$$\bar{k} = \left(\prod_{r=v}^{w} r \right)^{\frac{v}{v-w+1}}. \tag{2.19}$$

The output of this step is the tonal and noise maskers, $P_{TM}(k)$ and $P_{NM}(\bar{k})$ respectively, of each critical band.

3. There are two conditions used to remove unnecessary maskers. First, the maskers, obtained from the previous step which are lower than the absolute threshold of hearing $T_q(k)$ are removed. In other words, the survival maskers are those that satisfy the condition:

$$P_{TM,NM}(k) \geq T_q(k). \tag{2.20}$$

Second, within a distance of 0.5 Bark, the weak maskers are removed, and the strongest one is survived. Therefore, the output of this step is the relevant tonal and noise maskers of each critical band.

4. The masking contribution at the frequency bin i due to the tonal maskers located at bin j, $T_{TM}(i,j)$, is given by

$$T_{TM}(i,j) = P_{TM}(j) - 0.275 \times z_b(j) + SF(i,j) - 6.025, \tag{2.21}$$

where $P_{TM}(j)$ is the sound pressure level of the tonal masker in the frequency bin j, $z_b(j)$ is the Bark frequency of the bin j, and $SF(i,j)$ is the spread of masking from the masker bin j to the maskee bin i, defined as follows.

$$SF(i,j) = \begin{cases} 17\Delta_{z_b} - 0.4P_{TM}(j) + 11, & \text{for} \quad -3 \leq \Delta_{z_b} < -1, \\ (0.4P_{TM}(j) + 6)\,\Delta_{z_b}, & \text{for} \quad -1 \leq \Delta_{z_b} < 0, \\ -17\Delta_{z_b}, & \text{for} \quad 0 \leq \Delta_{z_b} < 1, \\ (0.15P_{TM}(j) - 17)\Delta_{z_b} - 0.15P_{TM}(j), & \text{for} \quad 1 \leq \Delta_{z_b} < 8, \end{cases} \tag{2.22}$$

where $\Delta_{z_b} = z_b(i) - z_b(j)$.

The masking contribution at the frequency bin i due to the noise masker located at the bin j, $T_{NM}(i,j)$, is given by

$$T_{NM}(i,j) = P_{NM}(j) - 0.175 \times z_b(j) + SF(i,j) - 2.025, \tag{2.23}$$

where $P_{NM}(j)$ is the sound pressure level of the noise masker in the frequency bin j, and $SF(i,j)$ is the spread of masking from the masker bin j to the maskee bin i, defined as follows.

$$SF(i,j) = \begin{cases} 17\Delta_{z_b} - 0.4P_{NM}(j) + 11, & \text{for} \quad -3 \leq \Delta_{z_b} < -1, \\ (0.4P_{NM}(j) + 6)\Delta_{z_b}, & \text{for} \quad -1 \leq \Delta_{z_b} < 0, \\ -17\Delta_{z_b}, & \text{for} \quad 0 \leq \Delta_{z_b} < 1, \\ (0.15P_{NM}(j) - 17)\Delta_{z_b} - 0.15P_{NM}(j), & \text{for} \quad 1 \leq \Delta_{z_b} < 8. \end{cases} \tag{2.24}$$

The output of this step is the masking contributions $T_{TM}(i,j)$ and $T_{NM}(i,j)$ due to each maskers.

5. Given the absolute threshold of hearing for the frequency bin i, $T_q(i)$, the global masking threshold $T_g(i)$ is defined as follows.

$$T_g(i) = 10\log\left(10^{0.1T_q(i)} + \sum_{l=1}^{L} 10^{0.1T_{TM}(i,l)} + \sum_{m=1}^{M} 10^{0.1T_{NM}(i,m)}\right), \tag{2.25}$$

where L and M are numbers of tonal and noise maskers, respectively. The output of this step is signal-to-mask ratio (SMR), which is defined as the difference between the SPL of the global masking threshold and the PSD of the analyzed signal. Figure 2.14 shows an example of the SMR (red line) of one frame.

2.5 Summary

This chapter presented the background of audio information hiding, its general applications, and some famous watermarking techniques. The SVD-based audio watermarking was reviewed and analyzed intensively. We classified all SVD-based methods into two frameworks and discussed that the first framework has the problem of false positive detection. The advantages and disadvantages of SVD-based frameworks were also discussed.

All tools we used in our investigation and implementation were introduced, including the SSA, differential evolution, psychoacoustic principles, and the psychoacoustic model 1. We gave the interpretation of singular values which will play an important role in the following chapters.

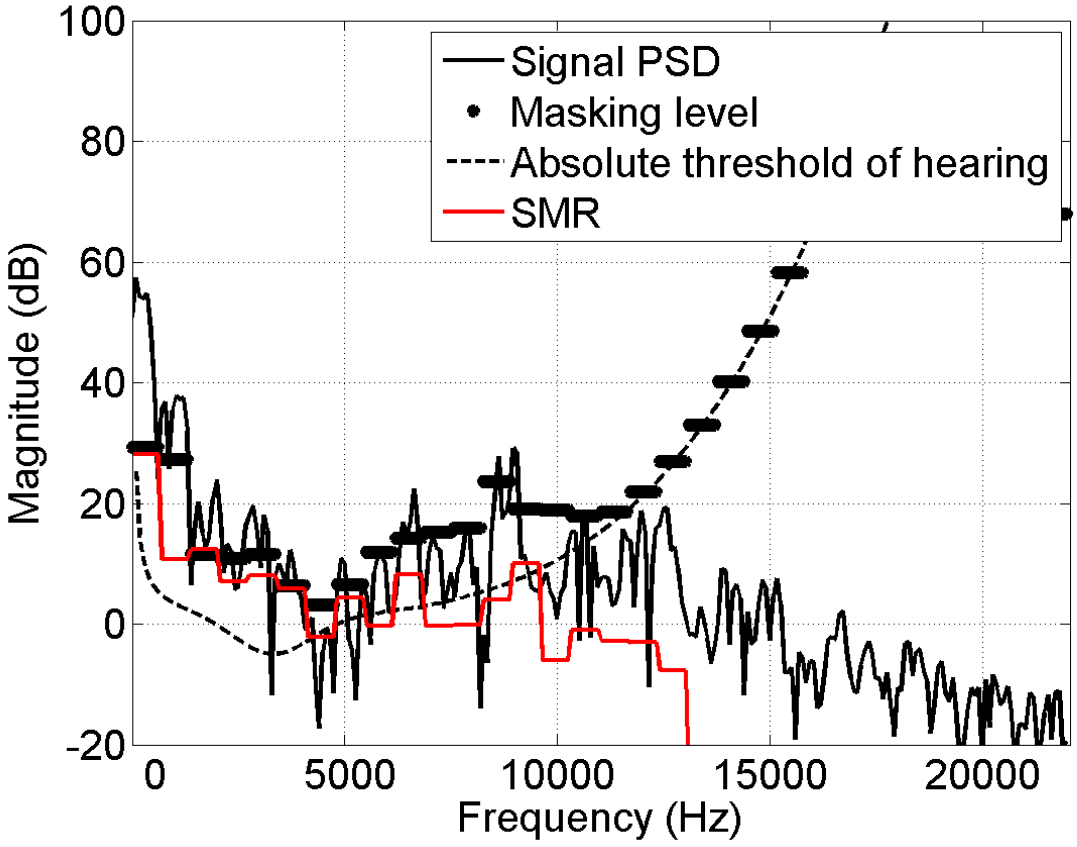

Figure 2.14: Signal PSD, global masking level, and SMR.

Chapter 3

Schemes of AIH based on SSA

SSA-based audio information hiding deploys the basic SSA, as described in the previous chapter, to analyze host audio signals. Fundamentally, the hidden information is embedded into the host signal by modifying some audio features with respect to a hidden-information bit. The details of how the modification can be made are provided in great detail in the following sections. We start with the core structure of the AIH based on SSA, on which the other improved schemes are based. The core is the simplest one, and in some contexts, it is equated with the *fixed-parameter model*. Thenceforth, the more complex schemes are introduced. We propose a novel automatic-frame-detection method and discuss the possibilities for AIH based on SSA in other domains in this chapter as well.

3.1 Philosophy of this work

Before we start to explain the proposed schemes in detail, we would like to emphasize the philosophy of this work, which makes this research different from other methods. This work aims to satisfy three requirements: inaudibility, robustness, and blindness. Satisfying these requirements has always been a difficult task. For example, the least-significant-bit coding technique [69–71] is good at the inaudibility, but it is not robust. The phase-manipulation techniques, such as the phase coding [73] and the phase modulation [115], are based on the fact that the human auditory system is not sensitive to relative phase change. Thus, both are inaudible, but, for the phase-coding technique, the

embedding capacity is low, and it is non-blind. The techniques based on echo hiding is simple, blind, and robust [75–78]. However, adding echoes has a number of constraints in inaudibility due to the sensitivity of the human auditory system. The spread-spectrum-based technique is good in robustness but is poor in inaudibility and capacity [116]. The wavelet-based technique, which embeds information in the high-frequency band, is good in capacity but poor in robustness [20]. The trade-off between the inaudibility and the robustness can be found in the methods based on adaptive phase modulation [117], the methods based on periodical phase shift [118], and the methods based on cochlear delay characteristics [17, 119, 120] as well. The SVD-based methods [18, 23, 25–43], as detailed in Sect. 2.1.4, are robust but not good in inaudibility. These methods seem to be good in one property but not good in other properties. Thus, these examples show that it is very difficult to satisfy the inaudibility, the robustness, and the blindness, simultaneously.

To solve these problems of conflicting requirements, the SSA-based AIH framework is proposed. The philosophy of this work is to exploit the robustness from the SVD-based method and to improve its inaudibility by combining it with the human audio-perceptual model. This combination is possible because the SSA is an SVD-based analysis technique of which the singular values can be interpreted and have the physical meanings. Therefore, we can make a connection between the SSA-based framework and the human audio-perceptual model. We believe that the combination between the SSA and the psychoacoustic model can help us to overcome the problem of conflicting requirements.

The following sections in this chapter show development of SSA-based AIH towards the combination, and implementations and evaluation results are given in the next chapter.

3.2 Core structure of the SSA-based AIH

This section presents the embedding and extraction processes of the core structure. They are fundamental to all other SSA-based schemes. We also discuss embedding locations, the effect of embedding the watermark into high- and low-order singular values, and the concept of embedding repetitions.

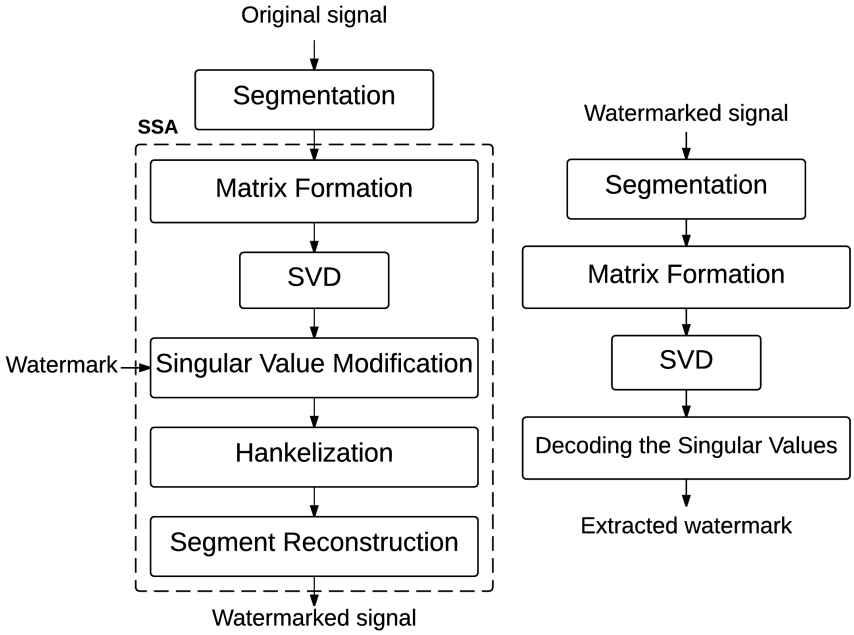

Figure 3.1: Embedding and extraction processes of the core structure

3.2.1 Embedding process

The embedding process consists of six steps, as illustrated in Fig. 3.1 (left).

1. Firstly, an audio signal is segmented into non-overlapping frames. The number of those frames is equal to that of hidden-information bits. For the simplest implementation, one bit of the hidden information (the watermark) is embedded into one frame. Thus, the embedding capacity is determined by the frame size.
2. Then, trajectory matrices representing each frame are constructed.
3. SVD operation is performed on each matrix. The steps 2 and 3 correspond to the embedding and SVD steps of the basic SSA algorithm, respectively. As a result, singular spectra are obtained.
4. A watermark bit is embedded by modifying the singular spectra.
5. After the singular-value modification, each modified trajectory matrix is Hankelized. This step corresponds to the diagonal averaging step of the basic SSA algorithm.
6. Finally, the watermarked signal is constructed by stacking those Hankelized frames.

The rule we use for the modification can be summarized as follows.

Let $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_d}\}$ denote a set of singular values in descending order and $\epsilon \in (0, 5]$ be a small real positive number. Given a watermark bit $w \in \{0, 1\}$,

$$\sqrt{\lambda_i} = \begin{cases} \sqrt{\lambda_l} + \epsilon(\sqrt{\lambda_u} - \sqrt{\lambda_l}), & \text{if } w = 0 \\ \sqrt{\lambda_l} + (1-\epsilon)(\sqrt{\lambda_u} - \sqrt{\lambda_l}), & \text{if } w = 1, \end{cases} \tag{3.1}$$

for $i = u+1, u+2, ..., l-1$, where $\sqrt{\lambda_u}$ is greater than $\sqrt{\lambda_l}$.

Figure 3.2 shows an example of a singular spectrum and its modification after embedding "1" and "0". Figure 3.2(a) is the singular spectrum. Given $u=20$, $l=50$, and $\epsilon=0.1$, Fig. 3.2(b) is its modification after embedding "1", i.e., $\sqrt{\lambda_{21}}, \sqrt{\lambda_{22}}, ...,$ and $\sqrt{\lambda_{49}}$ are set to $\sqrt{\lambda_{50}}+0.9(\sqrt{\lambda_{20}}-\sqrt{\lambda_{50}})$, and Fig. 3.2(c) shows its modification after embedding "0", i.e., $\sqrt{\lambda_{21}}, \sqrt{\lambda_{22}}, ...,$ and $\sqrt{\lambda_{49}}$ are set to $\sqrt{\lambda_{50}}+0.1(\sqrt{\lambda_{20}}-\sqrt{\lambda_{50}})$. Figure 3.3 shows an example of waveform when embedding "1" and "0" into the 35^{th} oscillatory component.

Note that we adopt just three of four steps from the basic SSA. The grouping step is neglect because its purpose is to separate a time series into meaningful additive sub-series, such as trends and noise, according to separability conditions [86]. Seeing that this is not the watermarking purpose, thus the step is excluded. However, one can consider that this is the case where the index m in Eq. 2.4 is equal to the index d in Eq. 2.3.

3.2.2 Embedding areas

As discussed in Sect. 2.2.2 about the interpretation of the singular value, the lower singular-value index, the more contribution to the signal. Therefore, if robustness is required, the lower the index, the better performance. On the other hand, if fragility is required, the higher, the better. However, modifying low-index singular values affect the sound quality of the watermarked signal.

Figures 3.4 and 3.5 illustrate this phenomenon and show that the LSD and SDR vary with embedding areas. The LSD and SDR are used to measure the sound quality of the watermarked signal. In the experiment for this illustration, the relation between the embedding position index (p_i) on the x-axis on the figures and the parameter u is that $u = 1 + 4 \times (p_i - 1)$, where $l = u + 10$. In addition, we also found that the sound quality of the watermarked signal slightly be affected by the window-length parameter (L) of the basic SSA, as demonstrated in Figs. 3.6 and 3.7. In these demonstrations, the relation

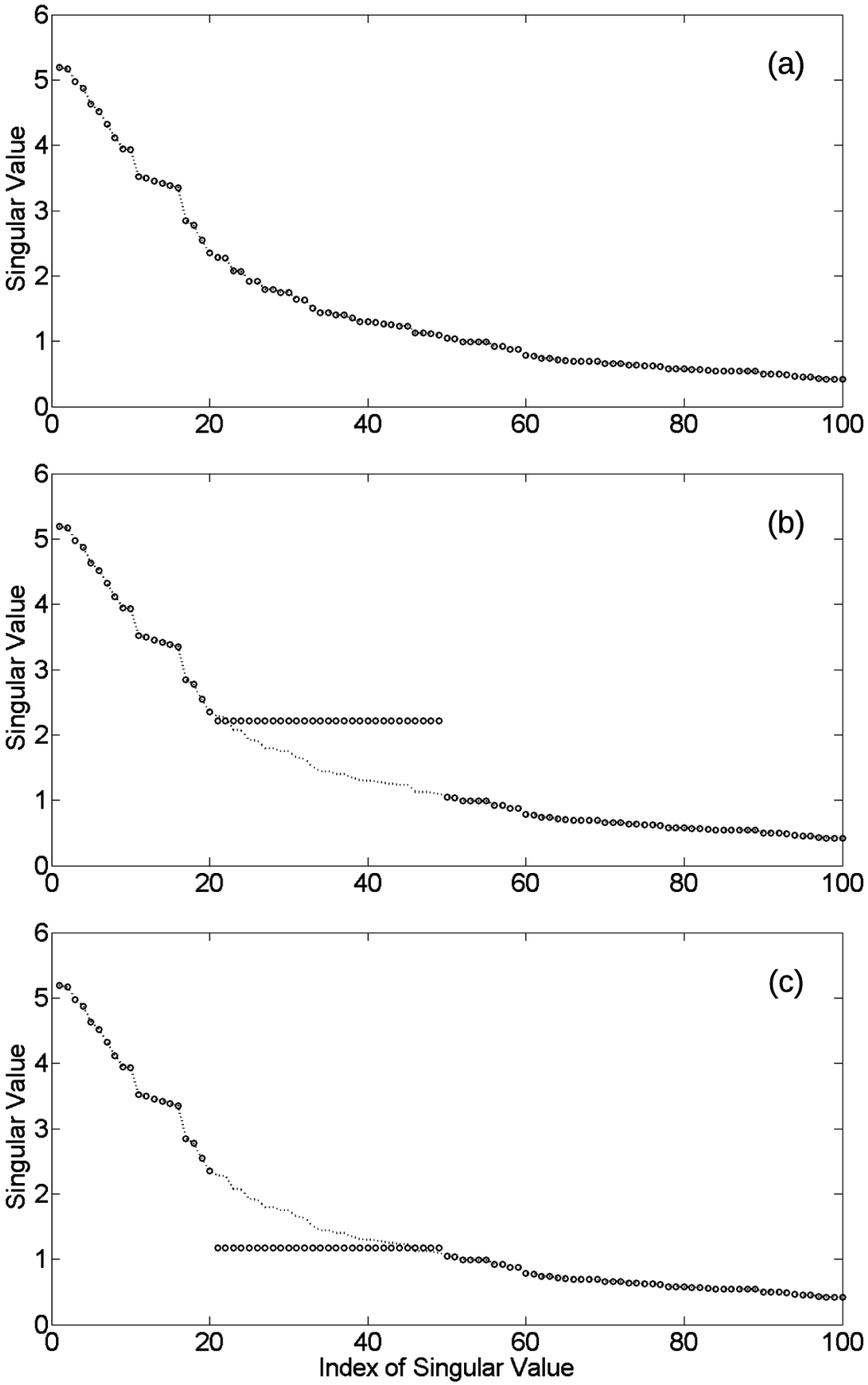

Figure 3.2: Illustration of the embedding rule of the core structure.

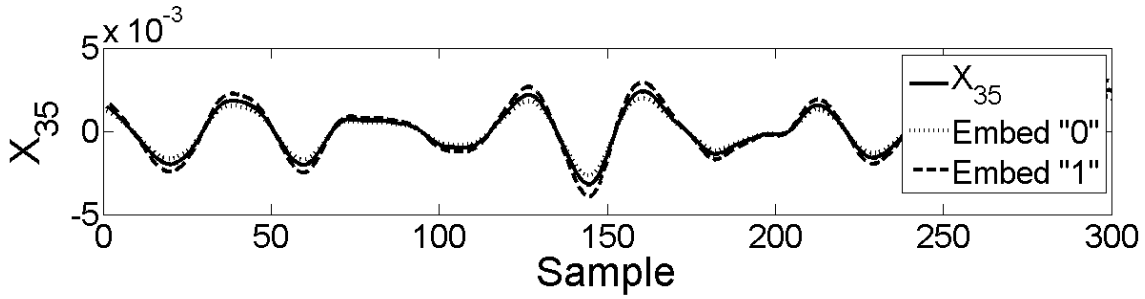

Figure 3.3: Example of embedding "0" and "1" into the 35$^{\text{th}}$ oscillatory component.

between the window-length index (w_i) on the x-axis and the window length (L) is that $L = \lfloor w_i \times N \rfloor$, where N is a frame length and $\lfloor \cdot \rfloor$ is the floor function.

Therefore, it seems that modifying high-order singular values is a good strategy. However, the higher-order singular values change easier than the lower-order ones when attacks are performed on the signal. Imitating the engineering strain (Cauchy strain), we define the strain e of a singular value $\sqrt{\lambda}$ as

$$e = \frac{\Delta\sqrt{\lambda}}{\sqrt{\lambda}}, \tag{3.2}$$

where $\Delta\sqrt{\lambda}$ is the change in the value of the singular value, i.e., the difference between the values of singular values before and after an attack.

We found that the strain increases as the order of the singular value increases. An example of this phenomenon is shown in Fig. 3.8, where a frame of 7350 samples is attacked by the MP3 and MP4 compression and the band-pass filtering.

3.2.3 Extraction process

The extraction process consists of four steps and is shown in Fig. 3.1 (right). The first three steps are exactly the same as those of the embedding process. The watermarked signal is segmented into several non-overlapping frames. Then, each frame is mapped to the trajectory matrix, and SVD is performed on the matrix to deliver the singular values. Lastly, the extracted watermark bits are decoded by using those singular values.

We have proposed two methods for decoding the watermark bit: decoding by the median and by the polynomial fitting.

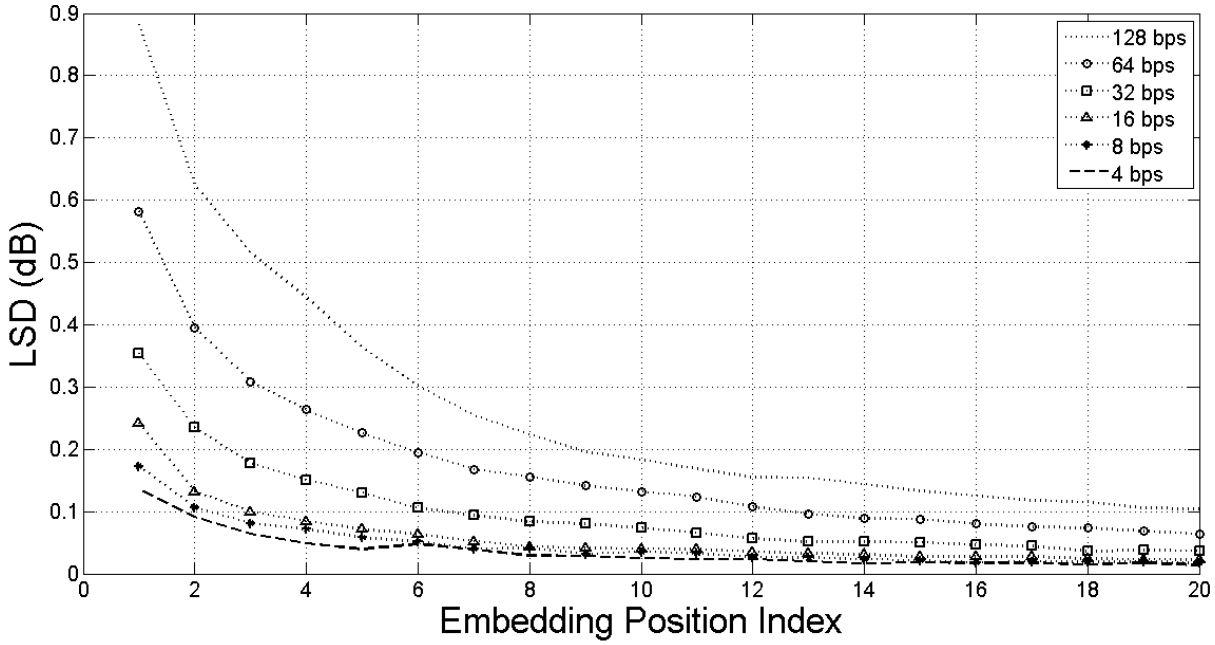

Figure 3.4: Relation between LSD and embedding area.

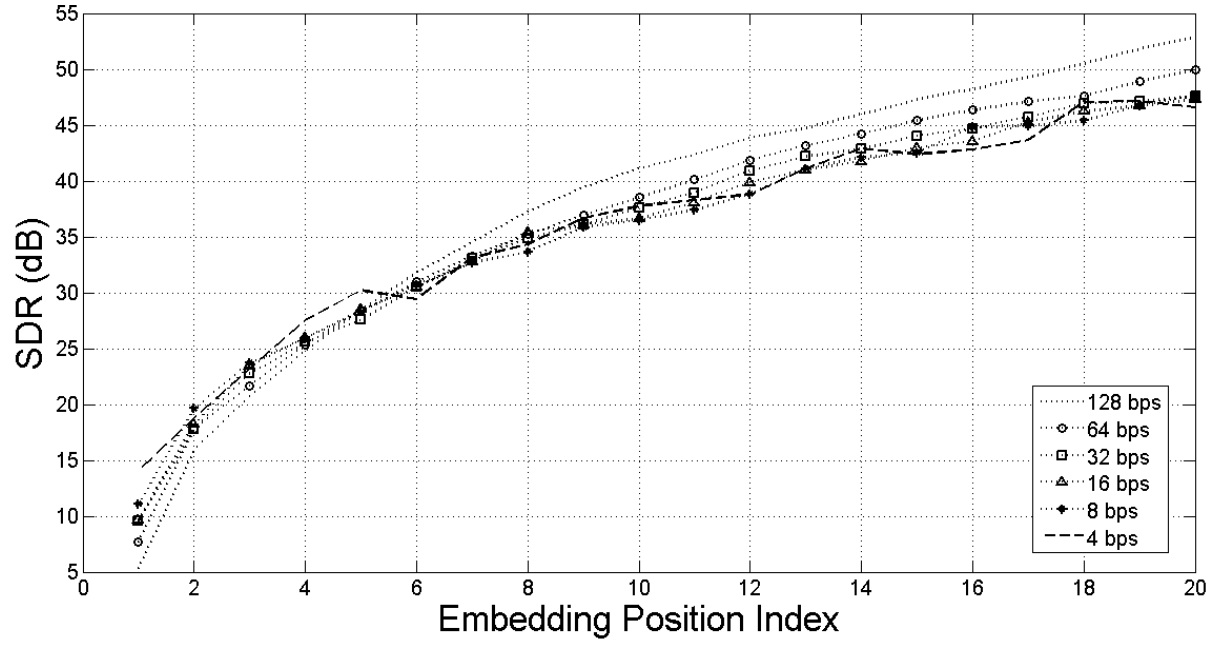

Figure 3.5: Relation between SDR and embedding area.

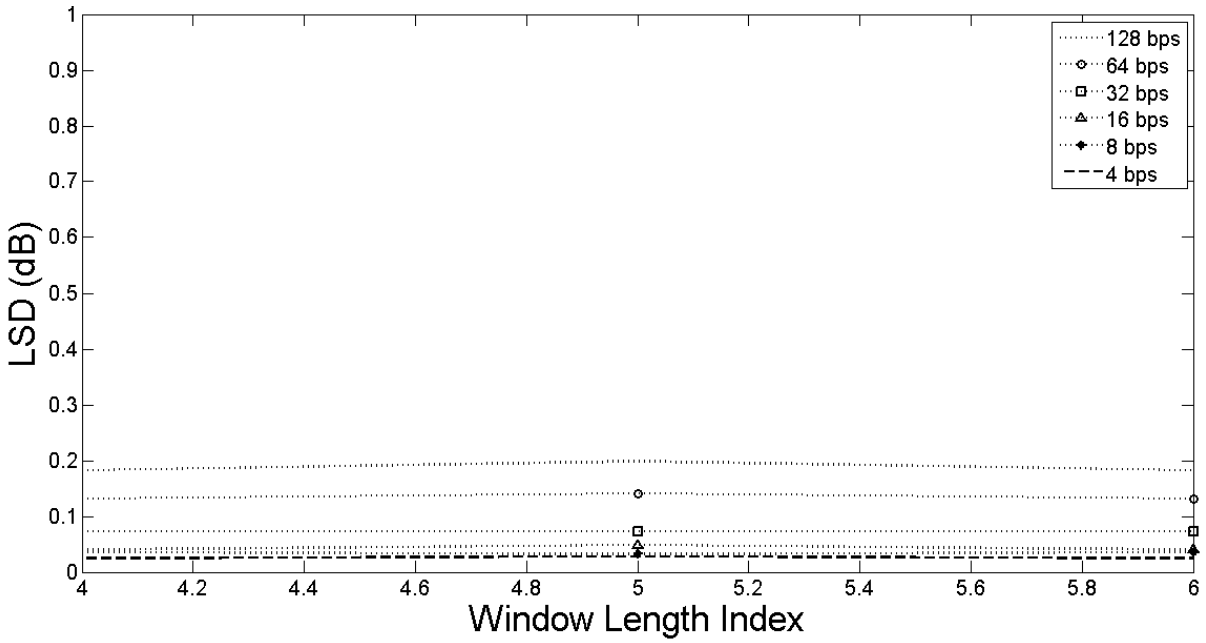

Figure 3.6: Relation between LSD and the window length (L).

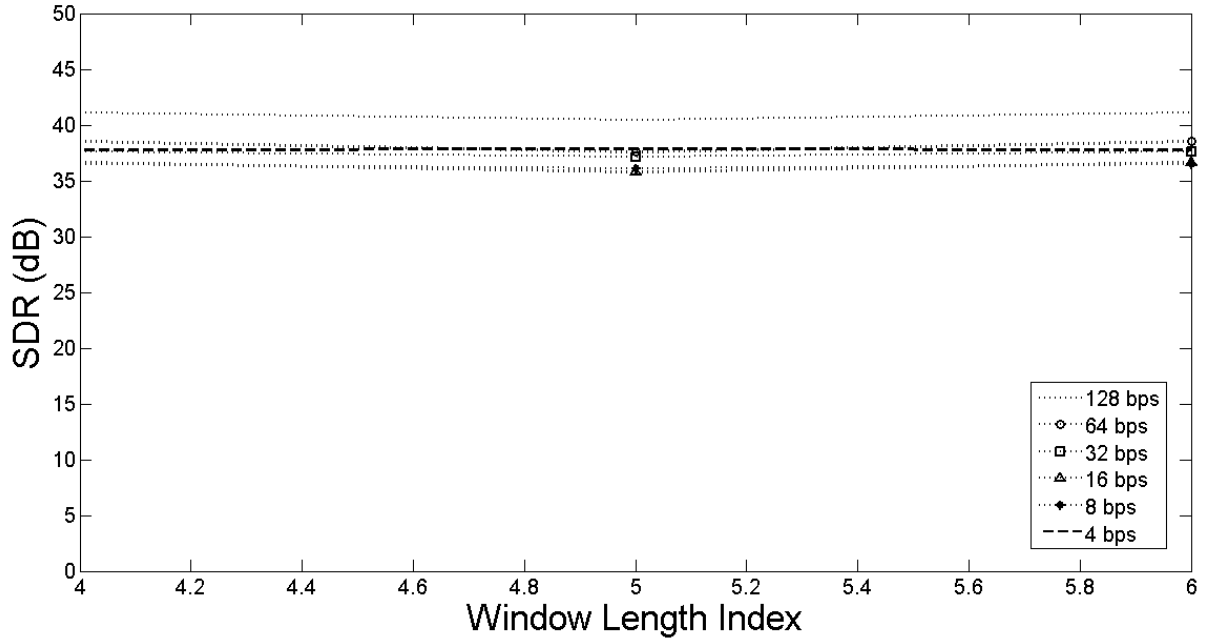

Figure 3.7: Relation between SDR and the window length (L).

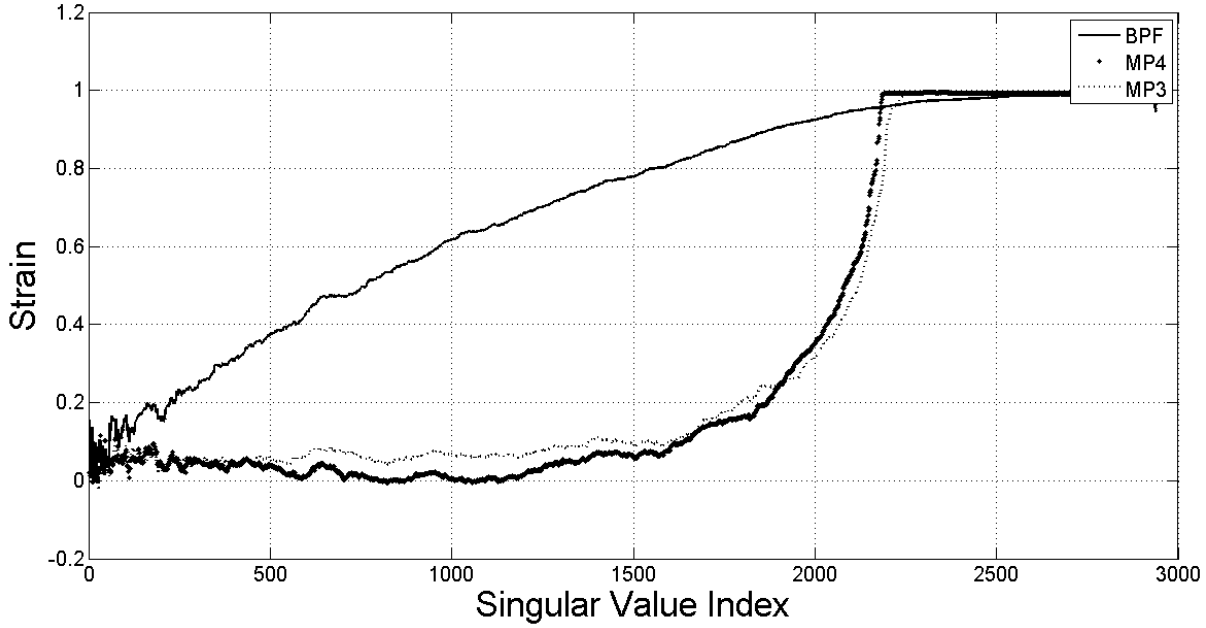

Figure 3.8: The ratio of total deformation of the singular value to the initial value, given the matrix size of 2940×4411.

Decoding by the median

The watermark bit is decoded by determining the value of $\sqrt{\lambda_m}$, where $\sqrt{\lambda_m}$ is the median of $\{\sqrt{\lambda_{u+1}}, \sqrt{\lambda_{u+2}}, ..., \sqrt{\lambda_{l-1}}\}$.

If $\sqrt{\lambda_u} - \sqrt{\lambda_m}$ is greater than $\sqrt{\lambda_m} - \sqrt{\lambda_l}$, the watermark bit is 1; otherwise, the watermark bit is 0.

Decoding by the polynomial fitting

An advantage of the median method is its simplicity in terms of both logic and computation. A drawback is that it uses only one singular value to decode the hidden bit, therefore it may raise a question of the reliability. By contrast, the decoding by the polynomial fitting uses all information of the singular values of $[u+1, l-1]$.

From the basic findings of this research, we found that when a singular spectrum of a reconstructed, watermarked frame is analyzed, the flatness resulting from rounding up or down of a sequence of singular values, according to the embedding rule, becomes distorted, as illustrated in Fig. 3.9. These distortions show patterns, as illustrated in Fig. 3.10, i.e., if singular values of $[u+1, l-1]$ are rounded down toward the lower bound, a convex upward curve is present; or else, a concave downward curve. The concavity or convexity of the

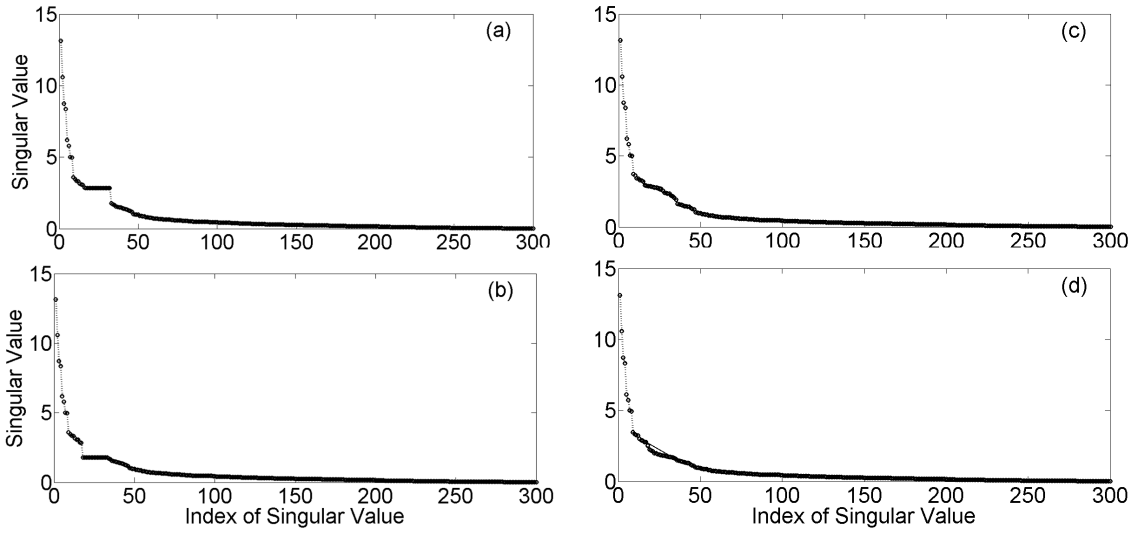

Figure 3.9: Example of singular spectra in embedding and detection processes in proposed method: (a) when embedding "1" and (b) when embedding "0" in the embedding process, (c) singular spectrum of the frame embedded "1" and (d) singular spectrum of the frame embedded "0" in the extracting process.

singular spectrum on a certain interval can be used to predict the watermark bit by the following algorithm.

All singular values of $[u+1, l-1]$ are fitted on a degree-two polynomial, $y(x) = ax^2 + bx+c$, where y is a singular value and x is the index of the singular value. The coefficient a of this quadratic formula has played an important role as it indicates the rate of change of the singular values. Therefore, a sign of the coefficient a is used to predict the watermark bit. That is, the minus sign indicates concavity or the watermark bit 1, and the plus sign indicates convexity or the watermark bit 0. An example of the polynomial fitting is shown in Fig. 3.11, where the watermark bit 1 is embedded into a frame by rounding the singular values of $[18, 32]$ up toward the upper-bound value, the 17^{th} singular value.

3.2.4 Embedding repetitions

Bit-detection rate (BDR) is defined as the number of correct extraction bits divided by the total number of watermark bits. A simple and straightforward way to increase the bit-detection rate is by embedding repetitions, i.e., embedding one watermark bit repeatedly into more than one subframes[1]. Figure 3.12 shows an example of embedding repetitions. One frame is partitioned into k subframes, where $k > 1$ is an odd integer, and

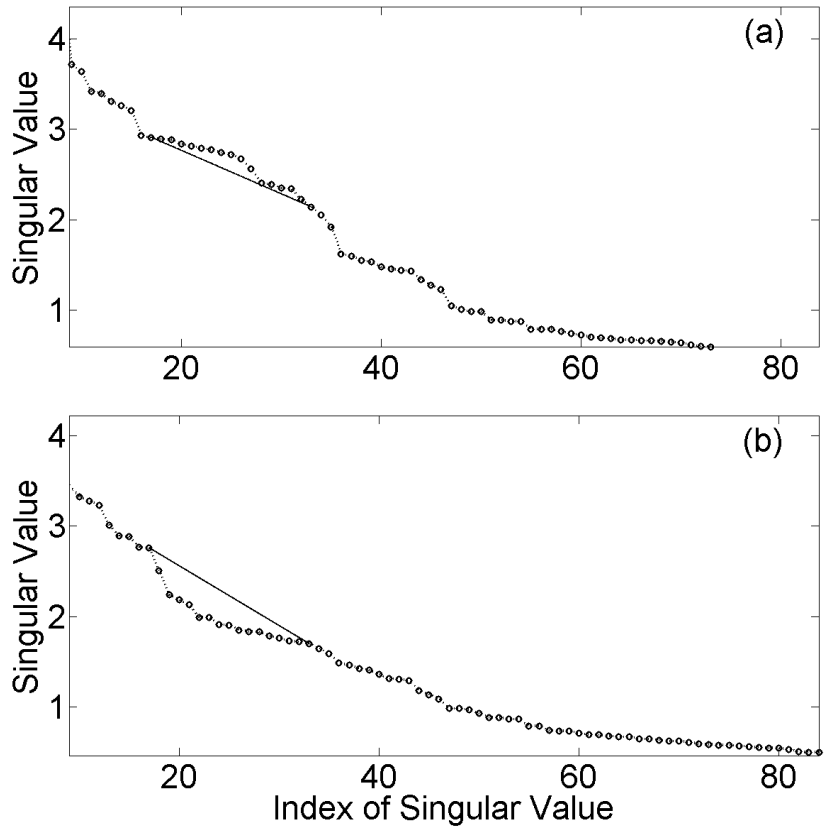

Figure 3.10: Distortion patterns in modified singular spectra: (a) On the interval $[u{+}1, l{-}1]$, the singular spectrum curve is concave downward if "1" is embedded. (b) The singular spectrum curve is convex upward if "0" is embedded. In this example, u and l are 17 and 33 respectively.

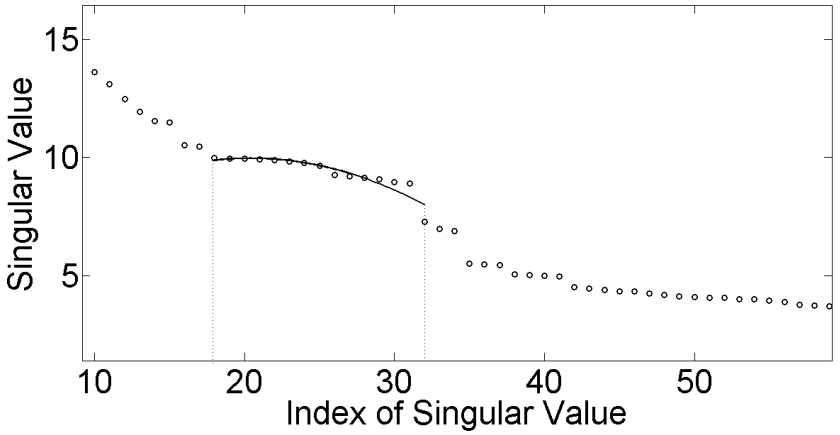

Figure 3.11: The singular values on $[18, 32]$ are fitted on a quadratic equation $y(x) = ax^2 + bx + c$, where $a = -0.0146$, $b = 0.598$, and $c = 3.853$. Since the value of a is negative, the graph is concave. Therefore, the watermark bit is 1.

all subframes of a frame are embedded the same watermark bit by the process described in Section 3.2.1.

To decode the watermark bit, the extraction process extracts all embedded bits from all subframes, and the majority rule was applied. The extracted watermark bit of the frame is 1 if $\lceil \frac{k}{2} \rceil$ of the extracted watermark bits of its subframes are 1, where $\lceil \cdot \rceil$ is the ceiling function; otherwise, the extracted watermark bit of this frame is 0.

The reason the embedding repetitions increases the bit-detection rate is very simple: if the BDR of a scheme without embedding a repetition is ϕ, then the BDR of the scheme when embedding k repetitions into k subframes of one frame, ϕ_k, is as follows.

$$\phi_k = \sum_{i=0}^{\lfloor \frac{k}{2} \rfloor} \binom{k}{i} (1-\phi)^i \phi^{k-i}, \tag{3.3}$$

where $\lfloor \cdot \rfloor$ is the floor function. It can be proved mathematically that if $\phi > 0.5$, then $\phi_k > \phi$. Figure 3.13 shows the relation between ϕ_k and ϕ when $k = 3, 7, 11, 15$, and 19. It can be seen clearly that when $\phi > 0.5$, all curves are above the line $\phi = \phi_k$.

3.3 Differential evolution and the SSA-based AIH

The core structure of the SSA-based audio information hiding, described in the previous section, has 3 parameters u, l, and ϵ. Since a singular spectrum from one audio signal is different from that from other ones, therefore it is logical to state that, to achieve a better result, those parameters should be adapted to the host audio signal. In other words, the SSA-based scheme can be improved by selecting the appropriate parameters, which is input-dependent, for embedding.

In this section, the cooperation between the core structure and the classic differential evolution is given in detail. The differential evolution is adopted as the parameter optimizer for three reasons:

- It is a multi-point optimizer. The effect of the starting point problem is mitigated, i.e., there is a good chance that the optimizer escapes the local minimum.

[1] To prevent confusions, the word *frame* is used in association with the embedding capacity, i.e., if the capacity is x bps, then within a duration of one second of a signal, there is x frames. A frame can be partitioned into many subframes for various purposed, and to increase the BDR is one of them.

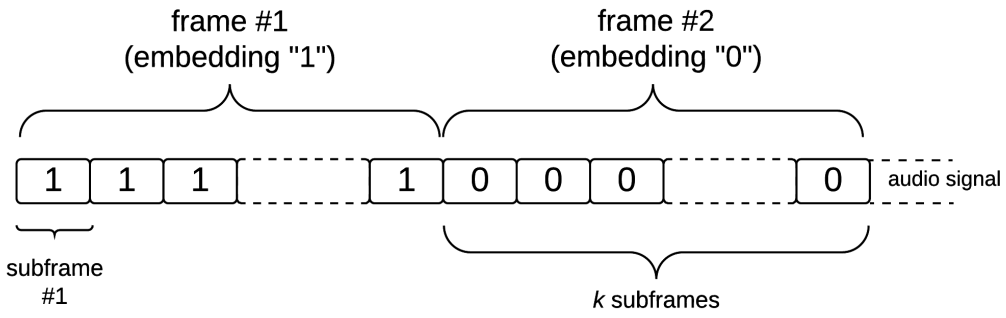

Figure 3.12: Example of embedding repetitions. Each frame is partitioned into several subframes, and all subframes of the same frame are embedded a same watermark bit.

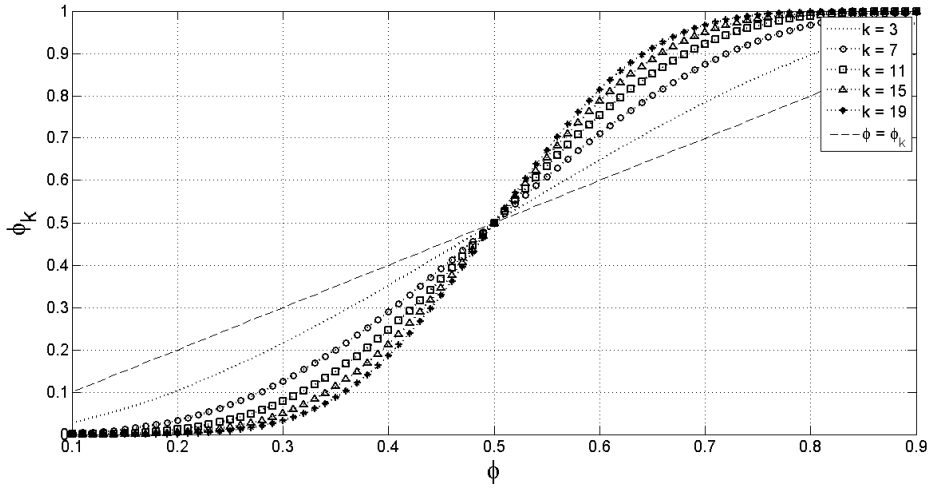

Figure 3.13: Bit-detection rate of the embedding repetitions.

- It is a derivative-free approach, which means we do not have to worry about whether the objective function is differentiable or not.
- It has good convergence properties and has proved to be the fastest algorithm in the evolutionary computational class [92].

3.3.1 Differential evolution optimization

Intuitively, the best parameters are those that yield the high robustness and introduce no perceptual distortion to the host audio signal. As mentioned in the first chapter, the two requirements conflict with each other. Thus, the optimization can be viewed as a trade-off mechanism. The optimizer deployed in this improved scheme aims to minimize the cost value defined as follows.

$$\text{Cost} = \sqrt{\alpha \cdot (\text{LSD} + (1 - \text{Sig}\,(\text{SDR})))^2 + \beta \cdot \overline{BER}^2}, \tag{3.4}$$

where LSD, Sig(SDR), and $\overline{BER}$ are the log-spectral distance, the sigmoid function of signal-to-distortion ratio (SDR) given that Sig$(\cdot)$ is a sigmoid function, and the average bit-error rate, respectively.

Given $P(\omega)$ and $\hat{P}(\omega)$ are power spectra of original and watermarked signals respectively, the LSD is defined as the following formula.

$$\text{LSD} = \sqrt{\frac{1}{2\pi} \int_{-\pi}^{\pi} \left[10 \log \frac{P(\omega)}{\hat{P}(\omega)} \right]^2 d\omega} \tag{3.5}$$

SDR is the power ratio between a signal and the distortion. Given amplitudes of original and watermarked signals, $A_{\text{org}}(n)$ and $A_{\text{wmk}}(n)$, the SDR is defined as follows.

$$\text{SDR} = 10 \log \frac{\sum_n [A_{\text{org}}(n)]^2}{\sum_n [A_{\text{org}}(n) - A_{\text{wmk}}(n)]^2} \tag{3.6}$$

Bit-error rate (BER) is defined as the number of error bits divided by the total number of embedded bits.

The term LSD + (1 − Sig(SDR)) in Eq. (3.4) represents a cost in terms of objective measures of the inaudibility, whereas the term $\overline{BER}$ represents a cost in terms of the robustness. The two user-defined constants α and β with relationship $\alpha + \beta = 1$ control the balance of inaudibility and robustness, respectively.

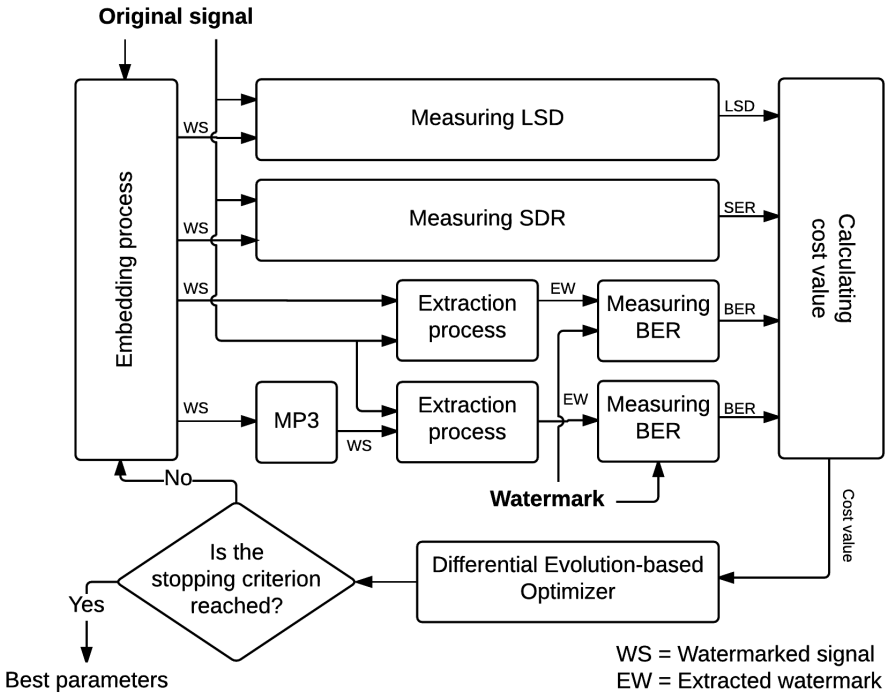

Figure 3.14: Differential evolution optimization.

The optimization is depicted in Fig. 3.14. Note that this is almost simplest because the optimizer simulates just one attack, i.e., the MP3 compression. For all practical or specific purposes, attacks can be simulated as many as needed so as to improve the performance of the scheme.

3.3.2 Automatic parameter estimation

Using the differential evolution to hand over the input-dependent parameters raises one problem: how to share those parameters between the embedding and extraction processes. To assume that the extraction process knows the parameters u and l in advance (fortunately, both extraction algorithms do not need the parameter ϵ) is not practical for some real-world applications. Therefore, a method for automatic parameter estimation is required. The new extraction process with automatic parameter estimation is shown in Fig. 3.15.

The automatic-parameter-estimation problem can be described as follows.

Let $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_d}\}$ denote a singular spectrum. According to the embedding

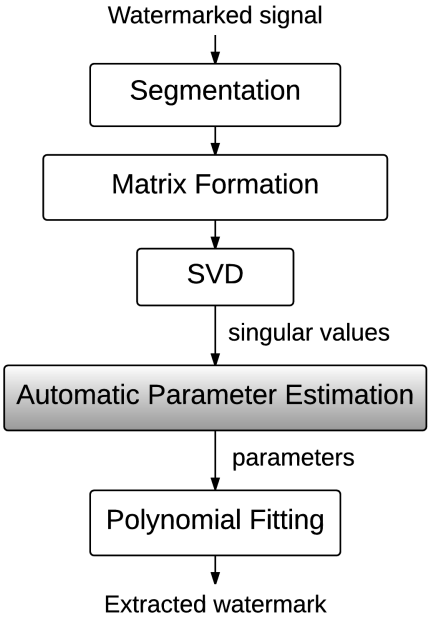

Figure 3.15: Extraction process with automatic parameter estimation.

rule, there are only two scenarios the set of singular values $\{\sqrt{\lambda_{u+1}}, \sqrt{\lambda_{u+2}}, ..., \sqrt{\lambda_{l-1}}\} \subset \{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_d}\}$, where $u < l < d$, is modified That is, it is either concave or convex.

Given a modified singular spectrum $\{\sqrt{\lambda_1}, ..., \sqrt{\lambda_d}\}$, where the $\{\sqrt{\lambda_{u+1}}, ..., \sqrt{\lambda_{l-1}}\} \subset \{\sqrt{\lambda_1}, ..., \sqrt{\lambda_d}\}$ is concave, and the possible maximum value of $l-u$ is known, the problem is to estimate parameters u and l. There are two reasons we are interested in only the concave scenario: First, naturally, a singular spectrum is convex, thus it is much easier to notice the concave part. Second, all frames in the whole audio signal share the same parameters, therefore, statistically, it is more than enough to estimate those parameters by using only the concave frames.

For the condition that the maximum possible value of $l-u$ is known in advance, this is not impractical. Because, although both parameters are input-dependent, it is not necessary that the maximum possible value of $l-u$ is input-dependent as well. In fact, the initialization process of the differential evolution requires that the search space is fixed, i.e., we predefine the possible minimum u and the possible maximum l. These two values are *input-independent* and determine the possible maximum $l-u$. Therefore, it can be shared between the embedding and extraction processes.

We have proposed two methods for automatic parameter estimation: the derivative-

based and concavity density-based methods.

Derivative-based method

Let Λ_i denote a singular value $\sqrt{\lambda_i}$. Given a singular spectrum $\{\Lambda_1, \Lambda_2, ..., \Lambda_d\}$, we define the first order derivative of the singular spectrum as $\{\Lambda'_1, \Lambda'_2, ..., \Lambda'_{d-1}\}$, where

$$\Lambda'_i = \Lambda_{i+1} - \Lambda_i, \tag{3.7}$$

for $i=1$ to $d-1$.

The second order derivative of the singular spectrum is defined as $\{\Lambda''_1, \Lambda''_2, ..., \Lambda''_{d-2}\}$, where

$$\Lambda''_i = \Lambda'_{i+1} - \Lambda'_i, \tag{3.8}$$

for $i=1$ to $d-2$.

Figures 3.16(b) and 3.16(c) illustrate an example of the first and second derivatives of the singular spectrum shown in Fig. 3.16(a), respectively. It can be seen clearly that when the singular spectrum is unmodified, the second-order derivative of the singular spectrum is similar to an underdamped harmonic oscillator. When the singular spectrum is modified and has a concave part, as shown in Fig. 3.17(a), it causes additional abrupt changes in the slope of the singular spectrum. Therefore, spikes, which are caused by those changes, present in the second derivative, as shown in Fig. 3.17(b). The parameters u and l are estimated by detecting the spike and calculating a distance between the spike and the point at which the oscillation stops.

Concavity density-based method

To understand the logic behind the concavity density-based method, let us first consider Fig. 3.18. A line segment of length $n-m$ connecting two singular values $\sqrt{\lambda_m}$ and $\sqrt{\lambda_n}$ is drawn, where m and n are indices, and n is greater than m. Note that $n-m$ is the length of its projection on the index axis. Because of the convexity of the singular spectrum, for any pairs of indices m and n, most singular values of $[m+1, n-1]$ are under the line segment.

However, when "1" is embedded, there exist some pairs of indices m and n such that the singular values of $[m, n]$ are mostly above the line segment, as shown in Fig. 3.19.

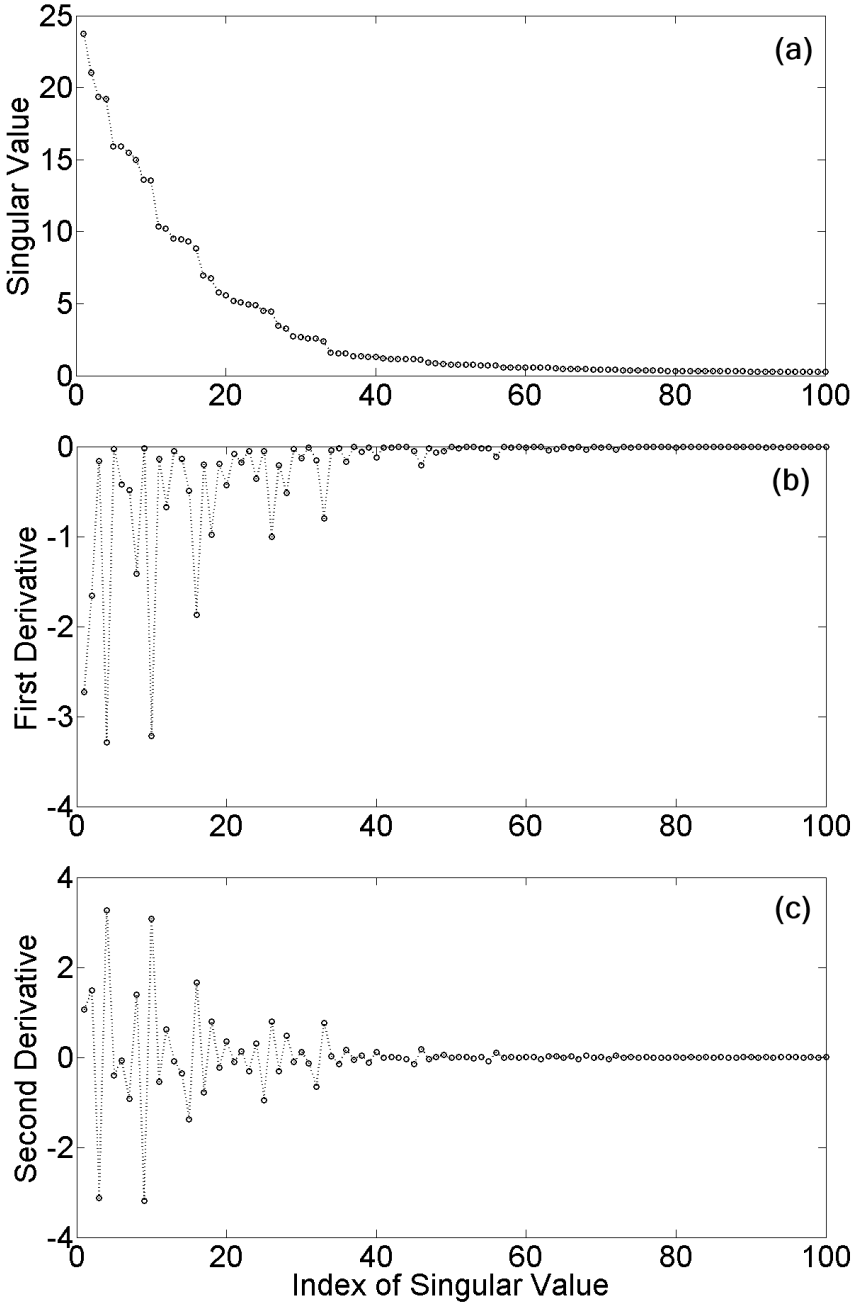

Figure 3.16: Example of a singular spectrum and its derivatives: (a) singular spectrum, (b) the first order derivative of (a), and (c) the second order derivative of (a).

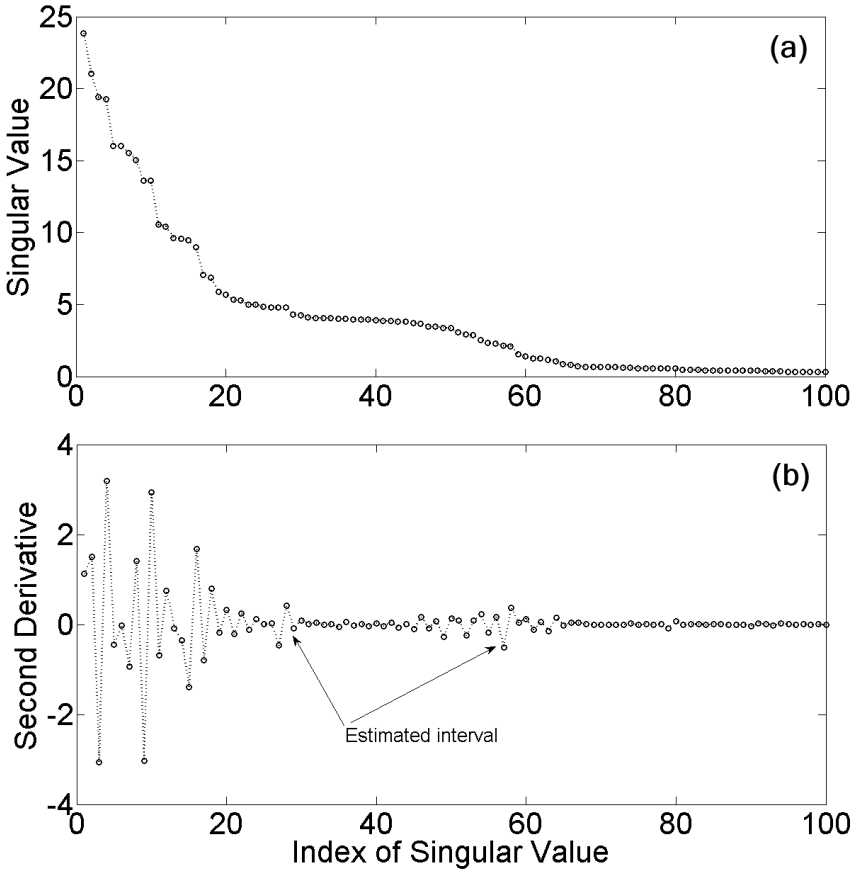

Figure 3.17: Example of a modified singular spectrum and its second order derivative: (a) modified singular spectrum with a concave part and (b) the second derivative of (a).

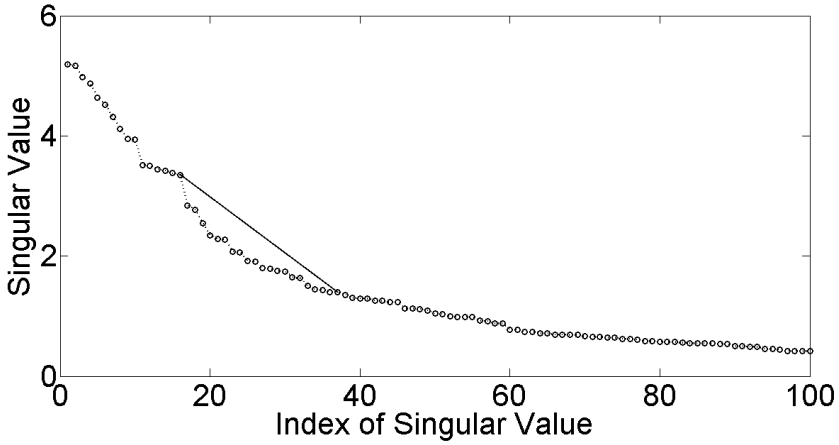

Figure 3.18: Singular spectrum of original signal and a line segment connecting $\sqrt{\lambda_{16}}$ and $\sqrt{\lambda_{37}}$. Singular values on $[17, 36]$ are under the line segment.

As a result, the number of singular values above given a line segment can be used as a clue to detect the concave part of the singular spectrum. This phenomenal is deployed to estimate the parameters u and l.

In this work, we define the concavity density as a measure of degree of the concavity. Given a singular spectrum $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_d}\}$, the concavity density $D_{m,n}$ of the singular values from $\sqrt{\lambda_m}$ to $\sqrt{\lambda_n}$ is defined as follows.

$$D_{m,n} = \sum_{i=m+1}^{n-1} \left(\sqrt{\lambda_i} - L(i) \right), \tag{3.9}$$

where $L(i)$ is the function defining the line connecting $\sqrt{\lambda_m}$ and $\sqrt{\lambda_n}$.

$$L(i) = \sqrt{\lambda_m} + \frac{\sqrt{\lambda_m} - \sqrt{\lambda_n}}{m - n} (i - m) \tag{3.10}$$

Basically, the $D_{m,n}$ is a sum of subtractions between singular values and the line connecting $\sqrt{\lambda_m}$ and $\sqrt{\lambda_n}$. An example of concavity-density calculation is demonstrated in Fig. 3.20. Starting from $m=1$, a line segment of length $n-m$, or a sequence of singular value that is used to calculate the concavity density, is shifted to the right one singular-value point at a time to determine a set of concavity density $\{D_{1,n}, ..., D_{i,n+i-1}, ..., D_{d-n+1,d}\}$ for $i=1$ to $d-n+1$.

Figure 3.21 shows an example of the concavity-density curve of the singular spectrum in Fig. 3.19 when the line segment has a length of 30. It can be seen from this example that the positive density corresponds roughly to the modification region of the singular spectrum. Hence, it is possible that the parameters u and l can be estimated by guidance from the concavity density.

The concavity density depends upon the choice of the length of a line segment, as shown in Fig. 3.22. If the length is too short, such as the line segments #1 and #2, or too long such as the line segment #3, we may not be able to detect the concavity. Figure 3.23(a) shows concavity-density curves with different lengths. Therefore, in order to correctly estimate the parameters u and l, we have to choose the appropriate length. In this work, we get around the problem by using the average density at different lengths. For example, the average density from Fig. 3.23(a) is shown in Fig. 3.23(b). To refine the density curve, two constraints are applied:

1. Any negative-density value is ignored because it implies convexity.

2. Any positive-density curve that is narrower than $\Gamma \cdot (l-u)$, where Γ is a user-defined real number around 1, is neglected because of the constraint on the minimum possible value of $l-u$ in the differential evolution optimizer.

Then, the indices at the rising and falling edges of the refined density curve, together with an offsetting constant, are used to estimate the parameters u and l for the given frame. Finally, the parameters u and l for the watermarked signal are calculated by averaging the estimated parameters u and l from all frames. The procedure is summarized as illustrated in Fig. 3.24.

Let $\hat{u}_{i,j}$ and $\hat{l}_{i,j}$ for $j = 1, 2, ..., k_i$ denote the estimated parameters of the frame i. The present of the subscript j indicates that it is possible that more than one concave intervals are detected within one frame. The maximum number of intervals detected within the frame i is denoted by k_i.

Given a set E of the estimated parameter-interval $[\hat{u}_{i,j}, \hat{l}_{i,j}]$, the averaging algorithm used to deliver the estimated parameters $\hat{u}$ and $\hat{l}$ is determined by the following procedure.

Given two integral intervals $[a, b]$ and $[c, d]$, where a, b, c, and d are integers and $b-a \leq d-c$, we say that there is an overlap between those two intervals if $c < b \leq d$ or $c \leq a < d$. For a pair of overlapping interval, $([a, b], [c, d])$, we define the overlap degree η as

$$\eta = \frac{\min(d-a, b-c)}{\max(b, d) - \min(a, c)}, \tag{3.11}$$

where $\max(\cdot)$ and $\min(\cdot)$ are the maximum and minimum functions, respectively.

As a consequence of the frame-parameter estimation algorithm, one thing we can expect from the set E is that it must contain a lot of overlapping intervals $[\hat{u}_{i,j}, \hat{l}_{i,j}]$. By the same token, we know that there is no overlap between intervals $[\hat{u}_{i,j_1}, \hat{l}_{i,j_1}]$ and $[\hat{u}_{i,j_2}, \hat{l}_{i,j_2}]$ when $j_1 \neq j_2$. Then, the averaging algorithm is just a process of recursively grouping the overlapping members of the set E, which consists of the following steps.

1. Each interval $[\hat{u}_{i,j_2}, \hat{l}_{i,j_2}]$ in the set E is assigned a number, called a frequency. Initially, all frequencies are assigned to 1.
2. Given a pair of estimated parameter-intervals, the overlap degree η is calculated. If η is greater than a predefined value, η^*, the two intervals are merged to create a new interval, i.e., the two intervals are removed from the set E, and the new one is

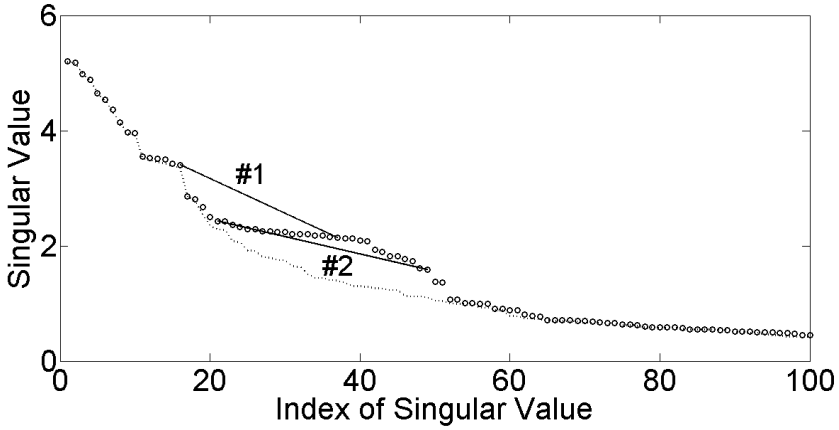

Figure 3.19: Singular spectrum of a watermarked signal and two line segments. Line segment #1 connects $\sqrt{\lambda_{16}}$ and $\sqrt{\lambda_{37}}$, and line segment #2 connects $\sqrt{\lambda_{21}}$ and $\sqrt{\lambda_{49}}$. Into this frame, a watermark bit 1 is embedded by forcing the singular values $\sqrt{\lambda_{21}}$ to $\sqrt{\lambda_{49}}$ toward the singular value $\sqrt{\lambda_{20}}$. The dotted curve represents the original singular spectrum. It can be seen clearly that almost all of the singular values are above the line segment #2.

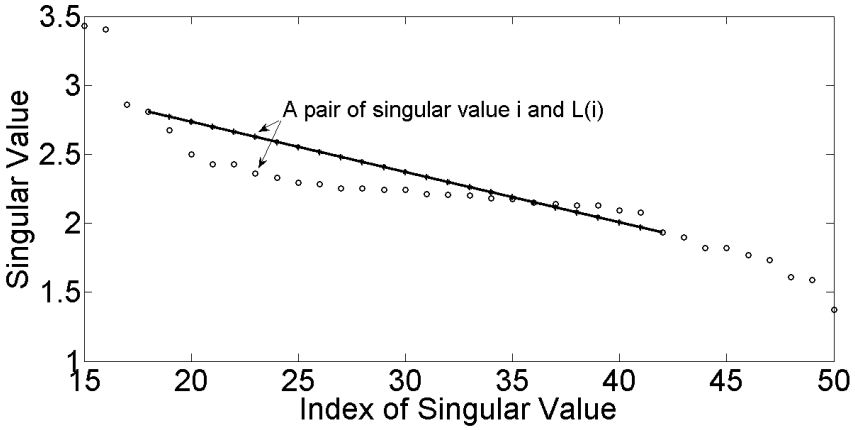

Figure 3.20: A line segment connecting $\sqrt{\lambda_{18}}$ and $\sqrt{\lambda_{42}}$ is used to calculate the concavity density $D_{18,42}$ by summing up all differences between singular values and their associated values on the line segment from index 19 to index 41. $D_{18,42}$ in this example is -2.5353. The minus sign implies that the segment of the singular spectrum on $[18, 42]$ is convex.

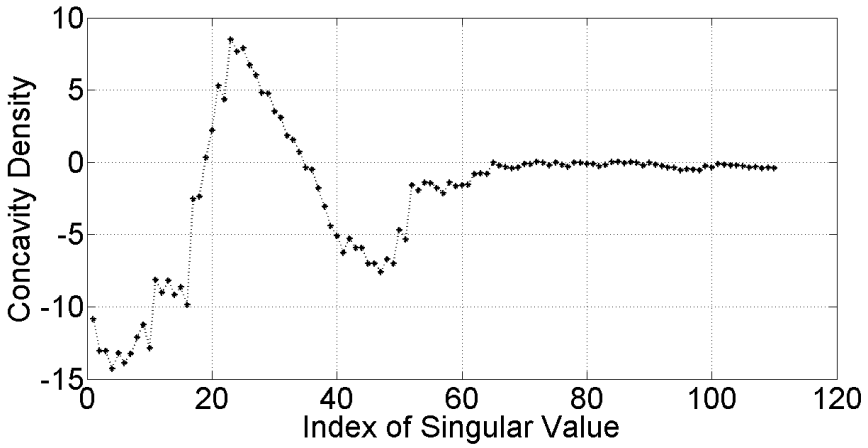

Figure 3.21: Set of concavity density $\{D_{1,31}, D_{2,32}, ..., D_{110,140}\}$. Notice that there is a strong relationship between regions of positive density and indices of modified singular values.

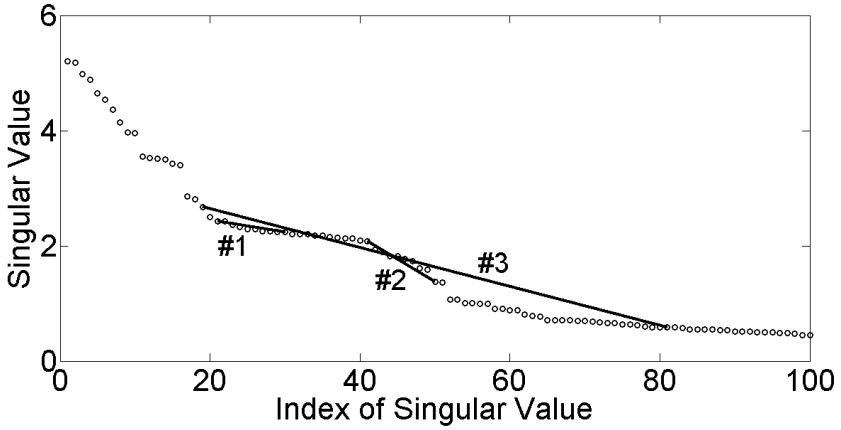

Figure 3.22: Singular spectrum as illustrated in Fig. 3.19 with three line segments. The line segment #1 connects $\sqrt{\lambda_{21}}$ and $\sqrt{\lambda_{30}}$, and the line segment #2 connects $\sqrt{\lambda_{41}}$ and $\sqrt{\lambda_{50}}$. They are too short. We cannot detect the concavity of the singular spectrum on $[21, 30]$ and on $[41, 50]$ even though singular values on such intervals are modified to embed "1". Line segment #3 connecting $\sqrt{\lambda_{19}}$ and $\sqrt{\lambda_{81}}$ is too long. It is difficult to obtain the positive concavity density from any long segment as well.

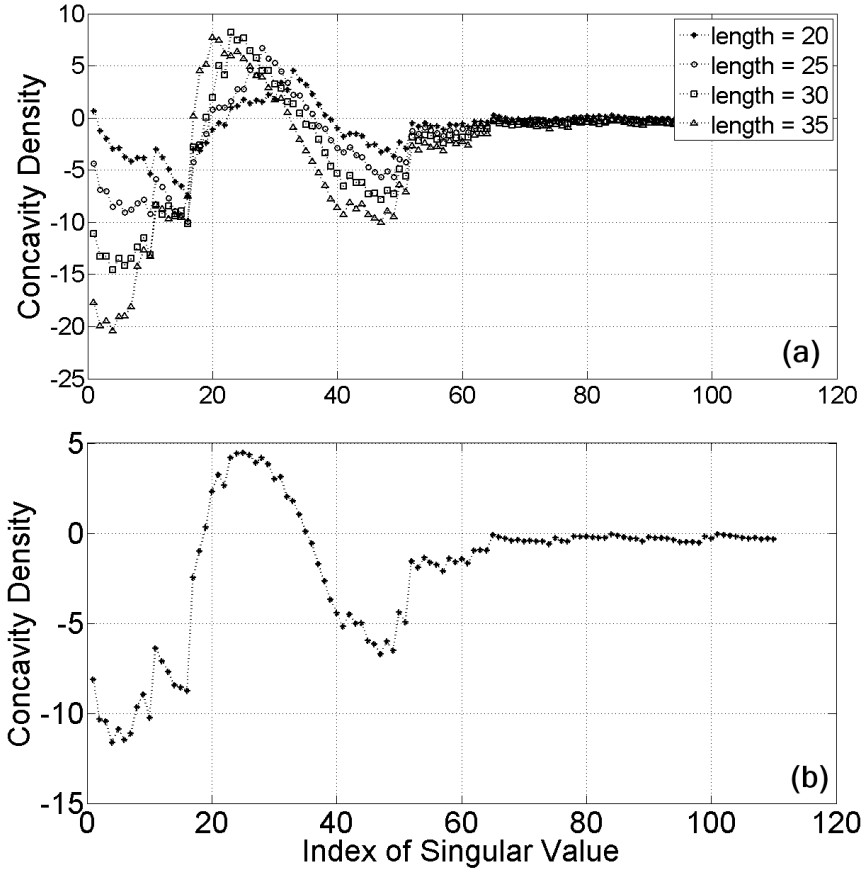

Figure 3.23: Concavity density curves when analyzing with four different lengths (a). An average of those four curves (b).

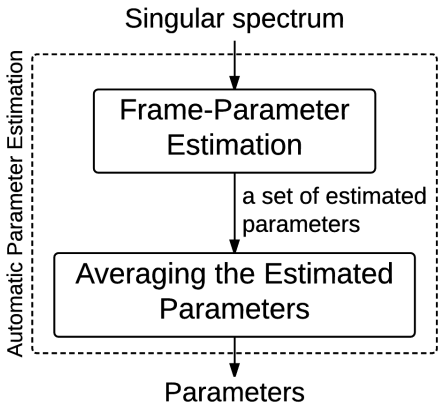

Figure 3.24: Automatic parameter estimation diagram.

added to the set. The frequency of the new interval is assigned by adding up the frequencies of the two old intervals. The new interval is $[\frac{a+c}{2}, \frac{b+d}{2}]$, where $[a, b]$ and $[c, d]$ are the two old intervals.

3. Step 2 is repeated until the set E has no overlapping member.
4. The interval with the highest frequency is chosen as the estimated parameters $\hat{u}$ and $\hat{l}$ for the watermarked signal. If there are more than one intervals with the highest frequency, the estimated parameters $\hat{u}$ and $\hat{l}$ are randomly chosen from them, uniformly.

The flow diagram of the algorithm is illustrated in Fig. 3.25.

3.4 Psychoacoustic model and the SSA-based AIH

The structure of this scheme is similar to the structure of the adaptive-parameter model. The only difference is that, instead of using the differential evolution optimizer to deliver the parameters u and l, this scheme asks the psychoacoustic model, as described in Sect. 2.4.4 to do that. The modified embedding process is shown in Fig. 3.26. The psychoacoustic model is implemented based on the psychoacoustic model 1.

The output of the psychoacoustic model is SMR. To use the SMR, we first have to establish the relation between frequencies and singular-value indices. As discussed in Sect. 2.2.2, singular values can be interpreted as scale factors of oscillatory components. When we use the Fourier transform to analyze each component, we can see that the frequency range of each component is narrow, as shown in Figs. 3.27 and 3.28.

In this work, we associate the index of singular value with the peak frequency of its oscillatory component. Then, the relation between the peak frequency of the oscillatory component and the index can be established. For example, the relation of a frame of 1024 samples is illustrated in Fig. 3.29. In this figure, we know for sure that the singular values from index 100 on have frequency components greater than 5 kHz. Therefore, given a frequency range, we can approximately map it to the singular-value interval.

The algorithm we use to deliver the parameters u and l is as follows. These steps are in the gray box of Fig. 3.26 and illustrated in Fig. 3.30.

1. The psychoacoustic model is used to calculate the SMR of each frame. According to

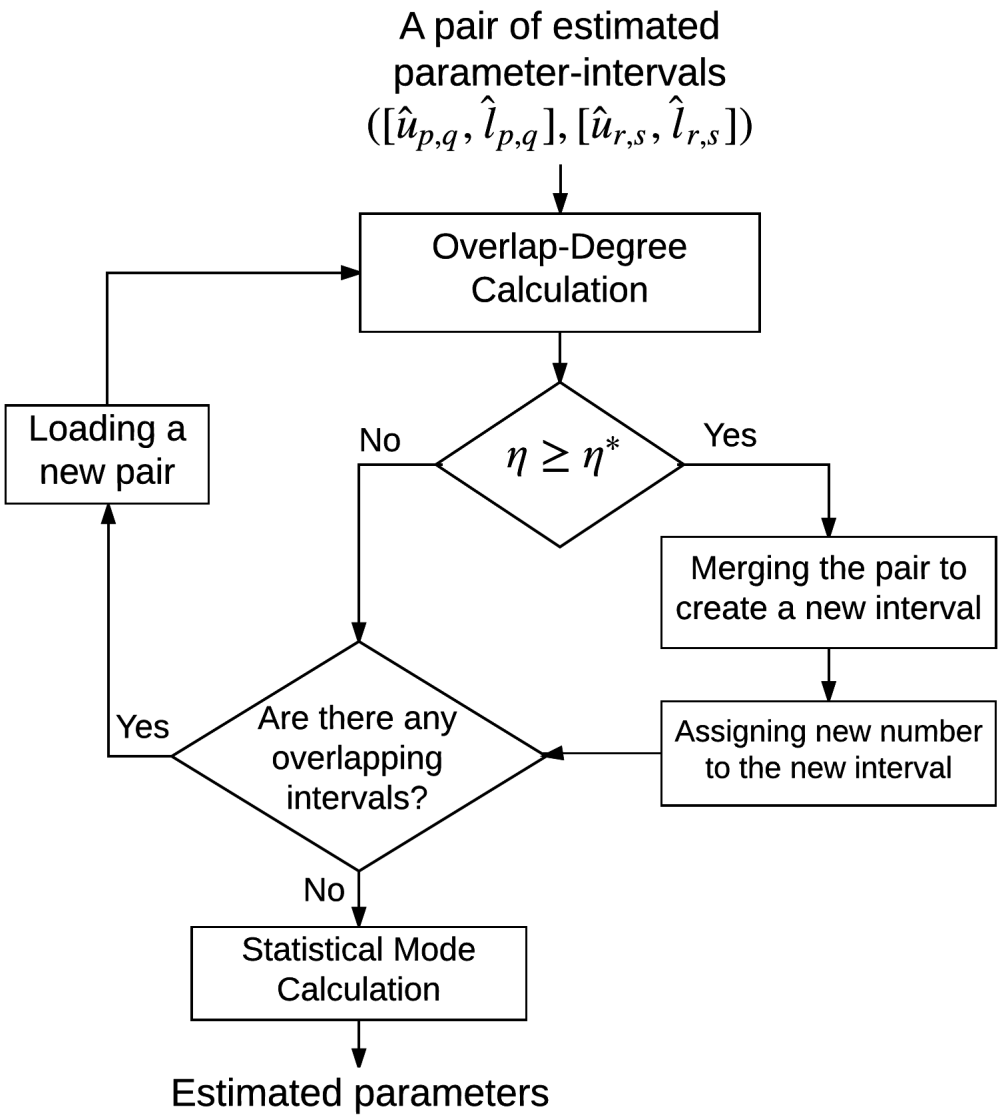

Figure 3.25: Averaging algorithm for the automatic parameter estimation.

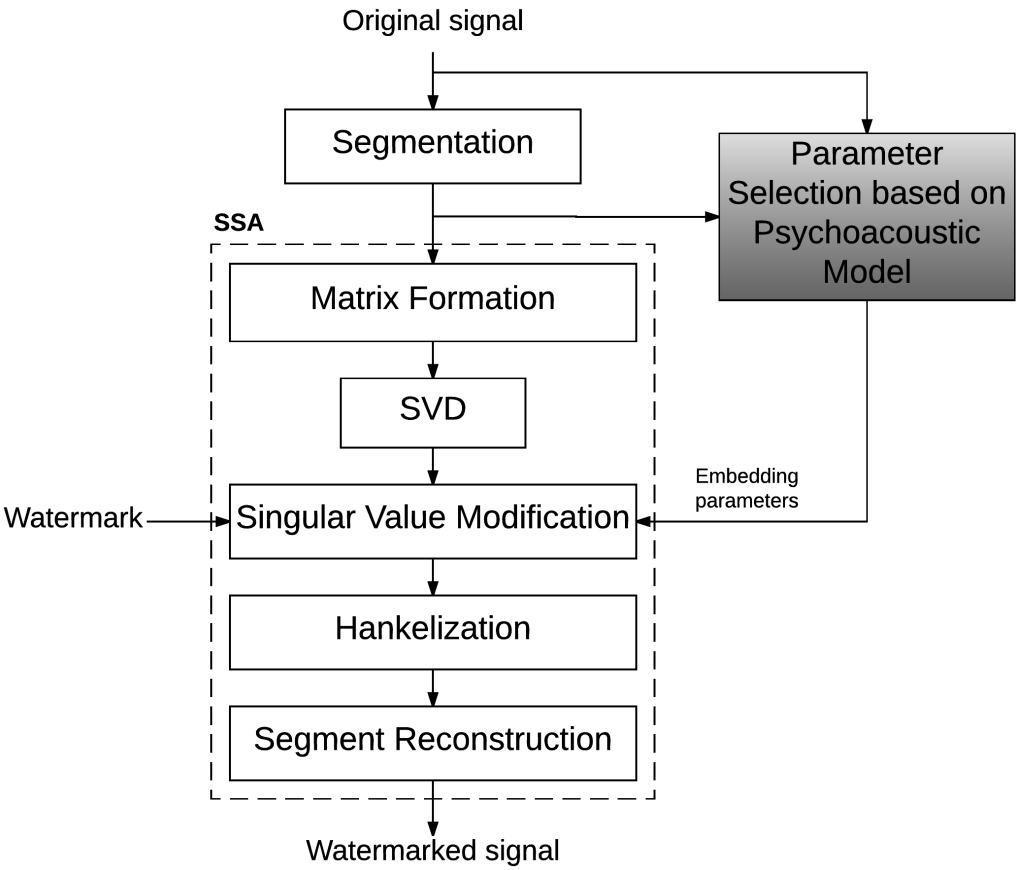

Figure 3.26: Embedding process of the AIH scheme based on SSA and a psychoacoustic model.

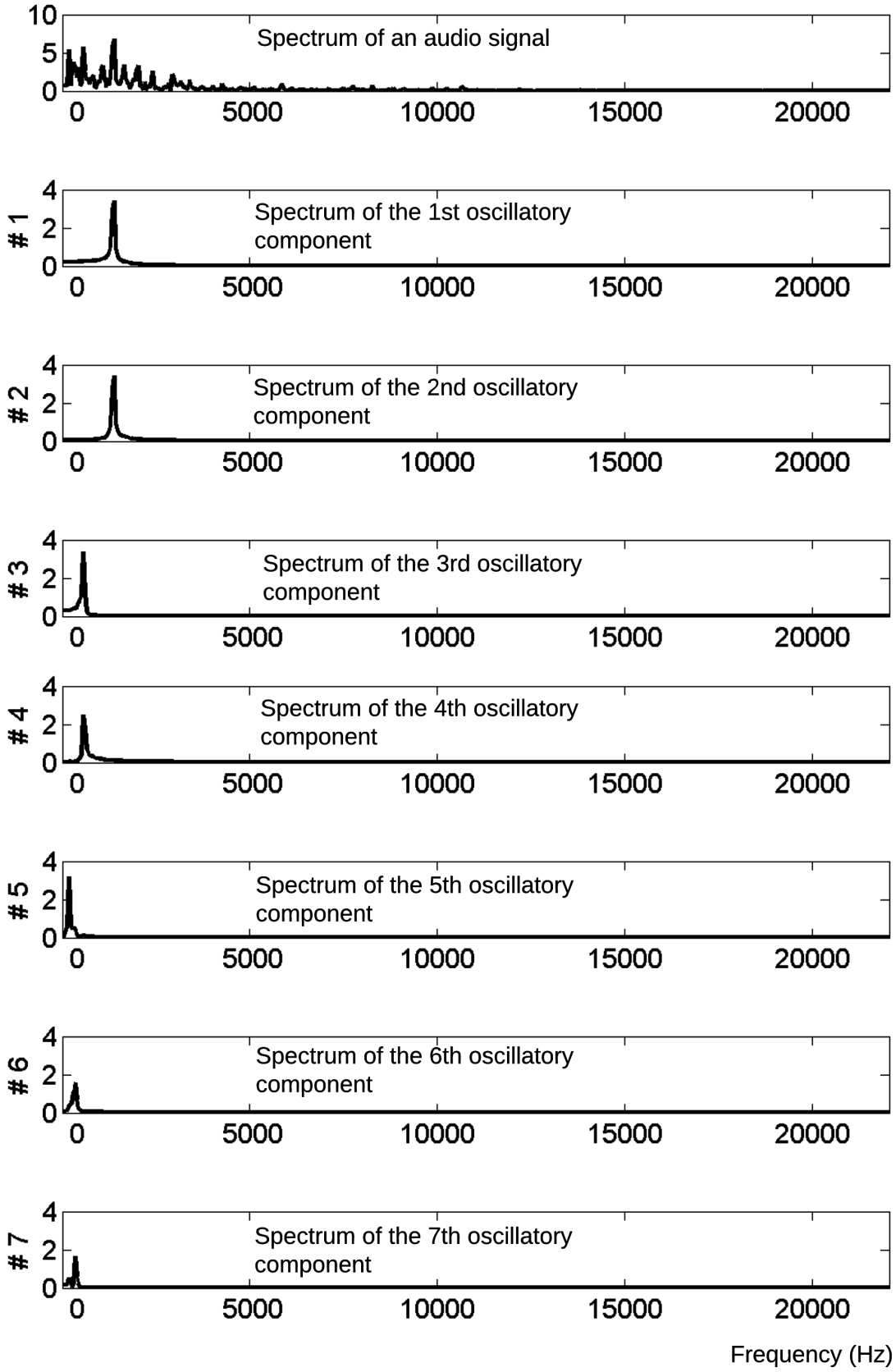

Figure 3.27: Example of the spectra of oscillatory components.

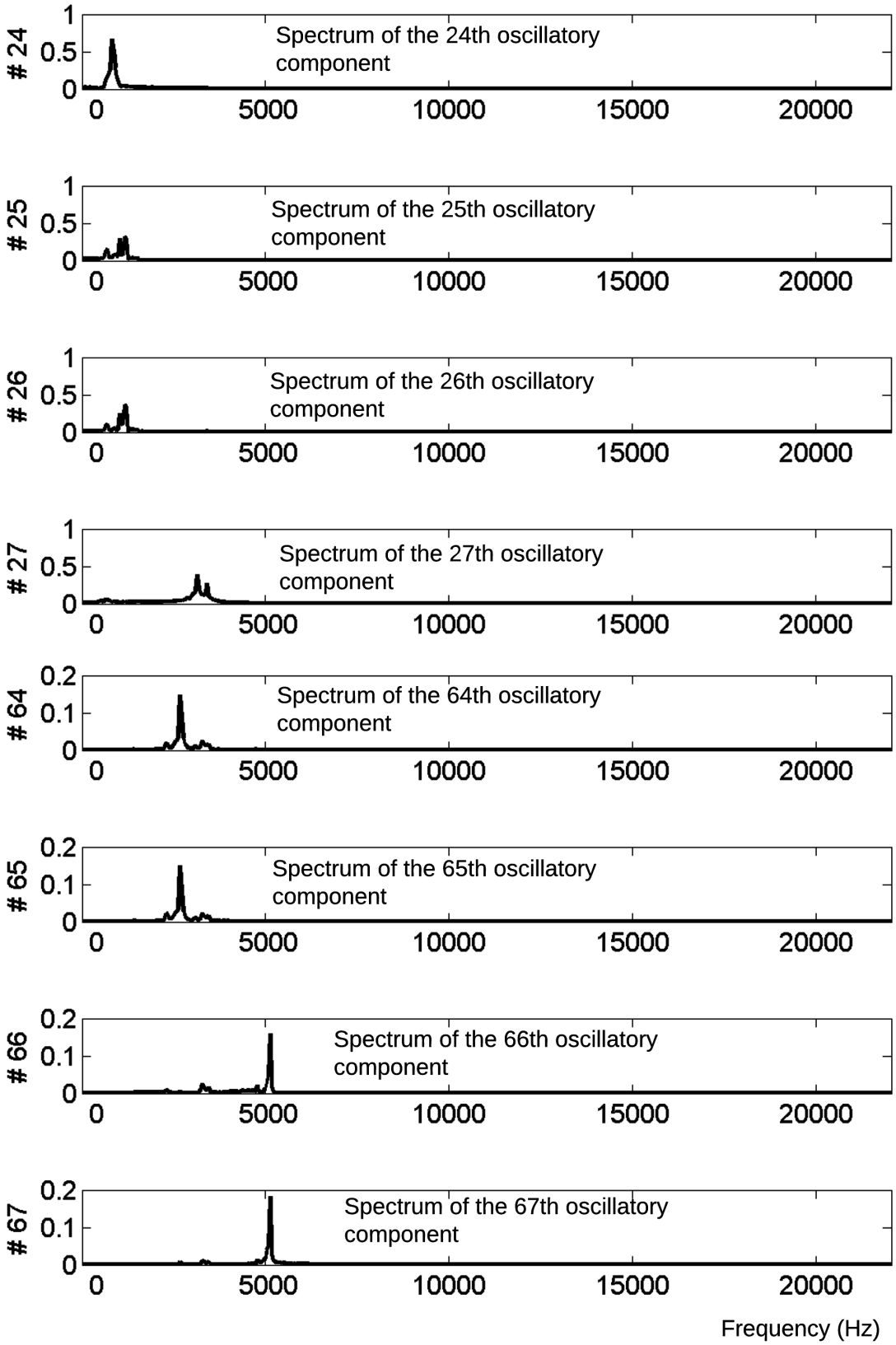

Figure 3.28: Example of the spectra of oscillatory components (continued).

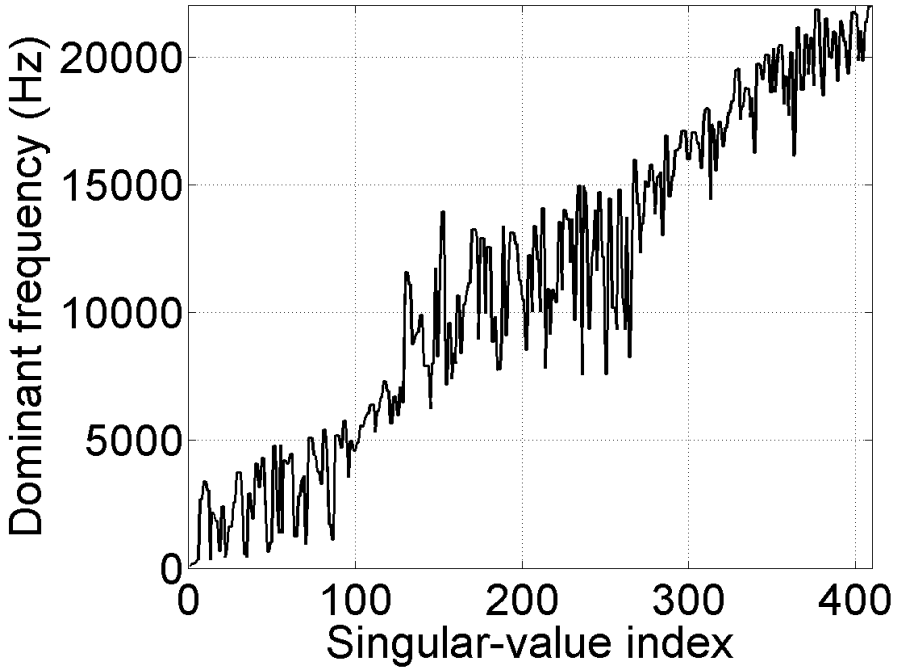

Figure 3.29: Relation between frequencies and singular-value indices.

the psychoacoustic model 1, the frame size used for this calculation is 512 samples.

2. The average SMR is calculated from all SMRs. An example of the average SMR of one frame is shown in Fig. 3.31.
3. We identify the frequency band $[f_1, f_2]$ with the average SMR lower than a predefined γ dB. If there are more than one band, the band with the lowest frequency is selected. If the frequency bandwidth $f_2 - f_1$ is wider than a predefined value, it is limited to that value. In our simulation, the predefined value is 10 kHz. An example is shown in Fig. 3.31.
4. The selected band $[f_1, f_2]$ is mapped to a singular-value interval $[u, l]$ for each frame. To map the frequency range $[f_1, f_2]$ to the interval $[u, l]$, we first find the local minimum which is closest to f_1, and u is the index at this local minimum. Then, we find the local maximum that is closest to f_2 which must be on the right side of u, and then l is the index at this local maximum. An example of this mapping is shown in Fig. 3.32.
5. Finally, the parameters u and l for embedding are the arithmetic mean of boundaries of all intervals.

After obtaining the parameters, the following procedure is exactly the same as that of the adaptive-parameter models.

3.5 Automatic frame detection for the SSA-based AIH

The embedding and extraction processes as described in previous sections are frame-based, i.e., the host signal is divided into frames, and one watermark bit is embedded into one frame. Actually, most proposed AIH schemes are frame-based. Thus, to correctly extract the watermark, the extraction process must know frame positions. The assumption that the extraction process knows frame positions in advance might not be hold in some practical situations. For example, an attacker can attack watermarked signals by cutting a few audio samples, known as the cropping attack. It causes the extraction process to work improperly. This problem is commonly called the frame synchronization problem.

There are two solutions to solve the frame synchronization problem [19]: (1) by binding the watermark with invariant audio features of host signal [121] or self-synchronization [122–124] and (2) by embedding a synchronization code into the host signal [125, 126].

3.5.1 Embedding synchronization code

Firstly, the most simplest techniques were studied. A synchronization code is embedded into every frame, where each frame is assumed to carry one bit of a watermark. A frame size is assumed to be constant. Let p_i denote the beginning position of a frame i, the goal is to find the set $\{p_1, p_2, p_3, ..., p_N\}$, where N is the total number of frames. Two methods were investigated and implemented [127].

LSB-based method

The LSB-based method generates a synchronization code by using normally distributed random number sequence and replaces the last L bits (the L right-most bits) of samples of each frame by a binary sequence derived from the generated random number sequence. Given an audio frame i of length M, the synchronization code embedding process consists of three steps as follows.

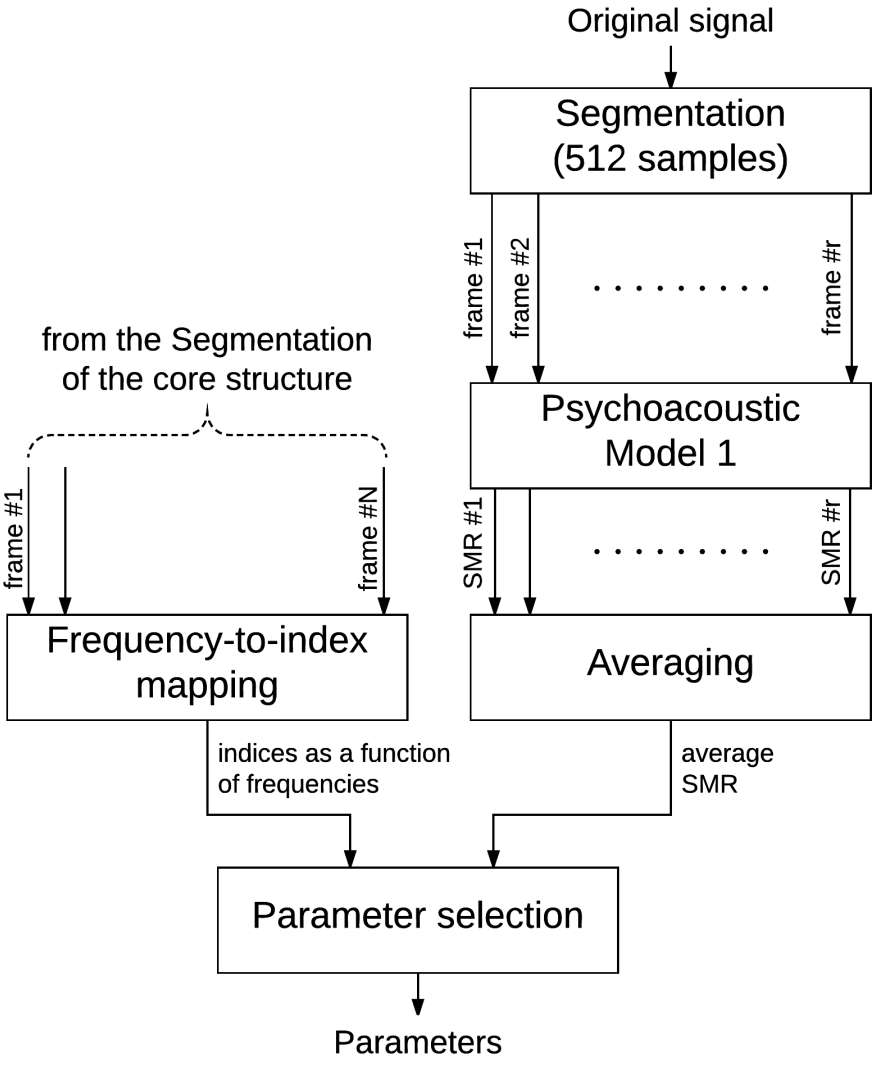

Figure 3.30: Parameter selection based on the psychoacoustic model 1. Note that the frame size from the segmentation of the core structure is not necessary to be the same as that of the psychoacoustic model.

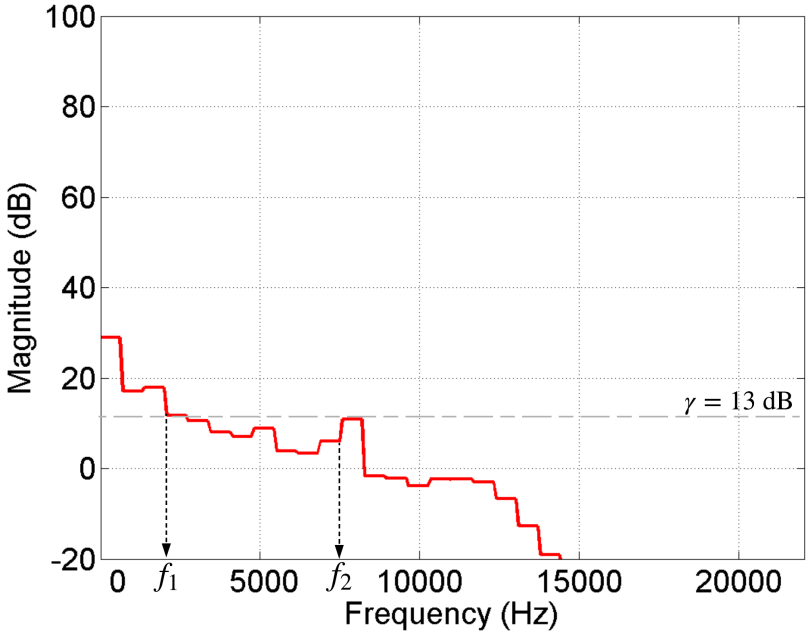

Figure 3.31: Average SMR.

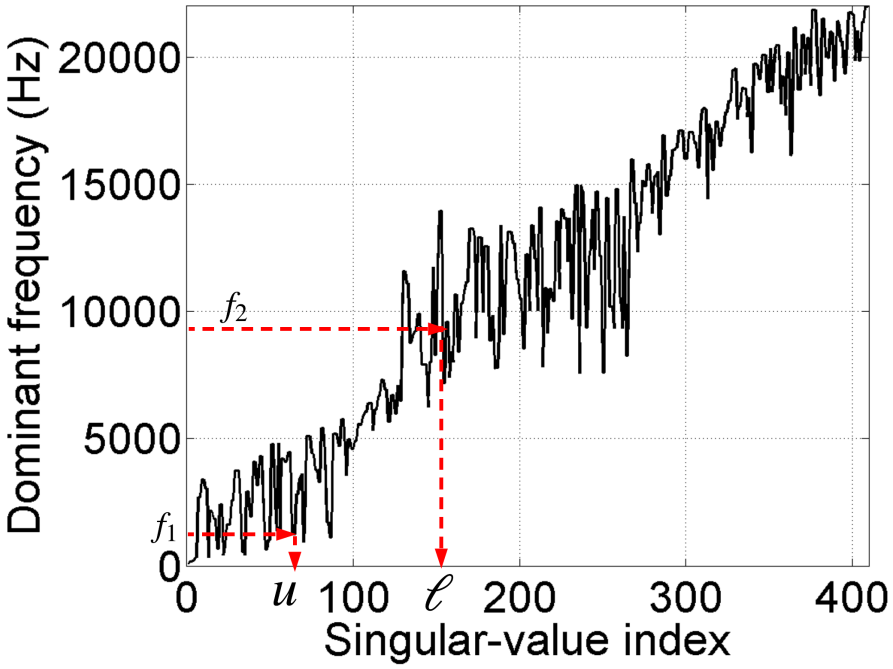

Figure 3.32: Example of the parameter selection: the frequency range $[f_1, f_2]$ is mapped to the interval $[u, l]$.

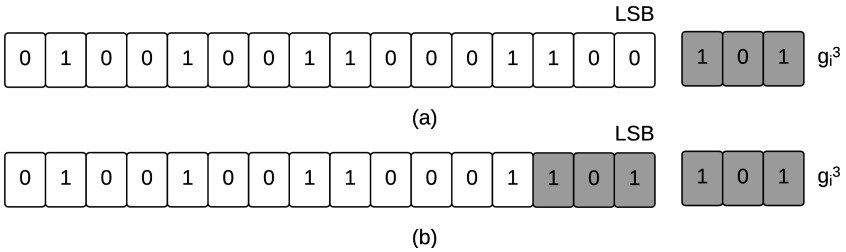

Figure 3.33: Example of replacing the last L bits of the sample i with g_i^L, where $L = 3$. (a) The last 3 bits of the sample are 1, 0, and 0, respectively, and g_i^3 is 101. (b) The last 3 bits of the sample are 1, 0, and 1, respectively, after the replacement.

1. Generate the normally distributed random number sequence $(g_1, g_2, g_3, ..., g_M)$ of lenght M.
2. Convert the decimal g_i to 16-bit binary number. If g_i^L is used to denote the last L bits of the binary number, the synchronization code is $(g_1^L, g_2^L, g_3^L, ..., g_M^L)$.
3. Replace the last L bits of the sample i with g_i^L. An example of this replacement process is shown in Fig. 3.33.

To detect the frame position, the cross-correlation between the sequence of last L bits of modified signal and the synchronization code is calculated. The frame position where the correlation is highest is detected as the correct position.

M-sequence method

M-sequence is a pseudo-random binary sequence generated by a linear feedback shift register (LFSR) generator [128]. One of the interesting properties is that the autocorrelation function of an M-sequence is very similar to a strain of Kronecker delta function. Thus, it is one of a good choice for the synchronization code. In this method, the M-sequence, $n(t)$, is added to each audio frame while keeping the signal-to-noise ratio (SNR) at a reasonable level. The synchronization code embedding process consists of three steps as follows.

1. Generate the M-sequence $n(t)$ of length M, where M is a frame size.
2. Determine a scale factor α for the M-sequence $n(t)$ in such a way that the SNR is

S dB. T_M is duration of frame length M in time.

$$\alpha = \sqrt{10^{-0.1S} \left(\frac{\int_{t=0}^{T_M} x^2(t)dt}{\int_{t=0}^{T_M} n^2(t)dt} \right)}, \tag{3.12}$$

where $x(t)$ is host signal.

3. Add the scaled M-sequence $\alpha n(t)$, which is the synchronization code, to the audio frame $x(t)$ directly. Hence, the audio frame with synchronization code is $x(t)+\alpha n(t)$.

To detect the frame position, beginning from the first sample, a frame of length N is selected, and the correlation between that frame and the synchronization code is calculated. Then, shifted to the right-hand side by one sample, the next frame of the same length is selected to calculate the correlation with the synchronization code again. This shifting process is continued until the last sample of the audio signal is reached. The correlation peak indicates the frame position.

From our preliminary experiments, we found that both methods could work well in the no-attack condition, but they are very fragile to signal processing attacks. We also found that the LSB-based method could achieve the acceptable sound-quality level, but it is not good enough when the fact that, in the simulation, no watermark information was embedded is taken into account [127].

For this reason, the proposed SSA-based AIH pursues the other solution to the frame synchronization problem.

3.5.2 Self-synchronization

We discover that, based on our proposed SSA-based audio watermarking, the scheme can also be used to detect watermarked frame automatically by analyzing singular spectral. However, to do this, we need to modify the embedding and extraction rules, and to fully grasp the idea behind the new rules, let us start with basic findings from this work.

Consider an audio signal with three equal frames of length M, where its middle frame is embedded the watermark bit 1 by the method described in Sect. 3.2.1, as illustrated in Fig. 3.34. The starting and the last indices of samples of the middle frame are denoted by n and $n+M-1$, respectively, and the frame is denoted by $G_{n,M}$. According to the

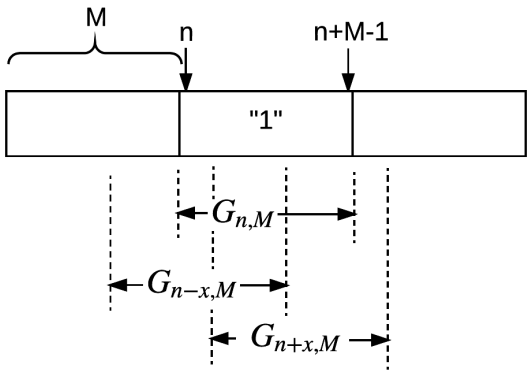

Figure 3.34: Example of the audio clip with 3 frames and three segments from which trajectory matrices are constructed.

embedding and extraction processes, if we use the frame $G_{n,M}$ to construct a matrix, then we can expect that there exists the concave part of the singular spectrum, as discussed in Sect. 3.3.2. If x is an integer less than M, then frames $G_{n,M}$ and $G_{n-x,M}$ are overlapping. We found that the singular spectrum of the trajectory matrix constructed by the frames $G_{n-x,M}$ also has a concave part if the overlapping area is large enough. The similar effect occurs to the frame $G_{n+x,M}$ as well. That is, if we construct matrices from frames $G_{i,M}$ for $i=0$ to $2M$, there are a lot of matrices that we can interpret as embedded the watermark bit 1. Those matrices are the ones that i is not too far from n. We call this the overlapping effect of embedding one, and it is the basis of the automatic frame detection. We name the process of constructing the frames $G_{i,M}$ for $i=0$ to the last possible frame and extracting the watermark bit from those frames the scan operation. This effect implies that we can detect the hidden bit one by performing the scan operation even though we cannot pinpoint the precise location of the watermarked frame. This is the reason we need the new embedding and extraction rules.

The new embedding procedure for the automatic frame detection is as follows.

We first divide a frame into 4 equal subframes. Suppose that each subframe has a length M. We represent one watermark bit by four bits of "0100" or "0110" depending upon the watermark bit. If the watermark bit is 0, four bits of "0100" are embedded into the 4 subframes. If the watermark bit is 1, "0110" are embedded into those subframes, as illustrated in Fig. 3.35. For example, if the watermark bits are "001", then the subframe-

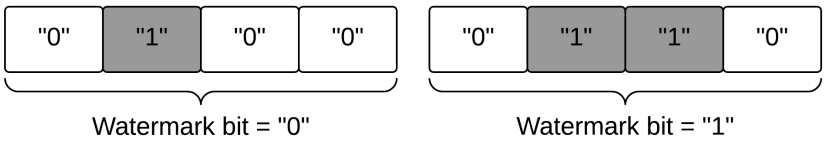

Figure 3.35: Four bits of "0100" are embedded into 4 subsegment of a frame, which represents embedding "0" (left), and 4 bits of "0110" are embedded into 4 subsegment of a frame, which represents embedding "1" (right).

embedding bits are "010001000110". In this scheme, embedding "1" into a subframe is the same as the embedding process of the core structure. However, embedding "0" into a subframe means leaving that subframe untouched. We do not need to force some singular values of the subframe toward the lower-bound value.

Given a frame $G = [g_0 g_1 ... g_{4M-1}]^{\mathrm{T}}$ of length $4M$, we define the subframe-scan operation on G as follows.

$$\mathrm{Scan}[G] = (b_0, b_1, ..., b_{\lfloor \frac{3M}{\delta} \rfloor}) \tag{3.13}$$

$$b_i = f(u, l, G_{i\delta,M}), \tag{3.14}$$

where δ is a scan step size, $G_{i\delta,M}$ is a subframe $[g_{i\delta}\, g_{i\delta+1} ... g_{i\delta+M-1)}]^{\mathrm{T}}$, for $i = 0$ to $3M$ of the frame G, and $f(u, l, G_{i\delta,M})$ is 1 if the singular spectrum curve of the matrix constructed from the subframe $G_{i\delta,M}$ on the interval $[u+1, l-1]$ is concave downward; otherwise, $f(u, l, G_{i\delta,M})$ is 0.

The meaning of this operation is that the scanner Scan$[\cdot]$, which operates on M samples, will scan through the frame G with the step size of δ and will return 0 or 1 depending upon the characteristics of the singular spectrum of the scanned subframe.

We use the first "1" of "0100" or "0110", or the first detection of concavity, as a synchronization point. If we can detect the next concavity, we will interpret it as the watermark bit 1. But if the concavity is detected for a short period, it is "0". Since we use the first detected concavity as a synchronization point, to make sure that all concavity are surrounded by convexity, and the distance between two concavity is far enough, "0" is added at the beginning and the end of the four-bit patterns. This is the reason behind using "0100" to represent the watermark bit 0 and "0110" to represent the watermark bit 1. An example of performing the subframe-scan operation according to Eq. (3.13) is shown in Fig. 3.36.

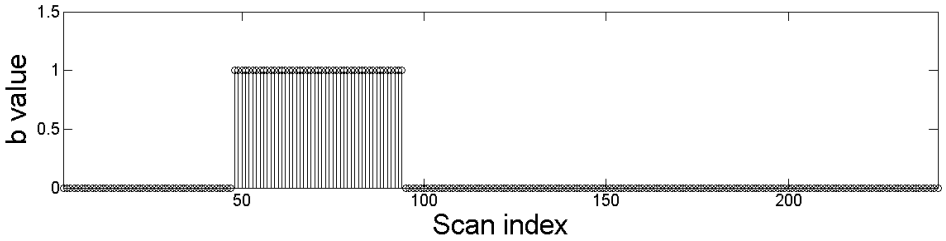

Figure 3.36: Example of performing the subframe-scan operation to a 3200-sample frame, i.e., $M = 800$, with $\delta = 10$, $u = 30$, $l = 80$, and "1" are embedded into the second subframe of the frame.

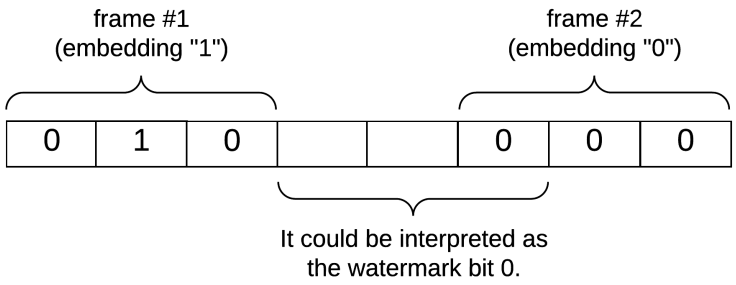

Figure 3.37: Three-bit patterns of "010" and "000" are used to represented the watermark bit 1 and 0, respectively. If unembedded frames are allowed, it is possible that we misinterpret some unembedded frames as having the watermark bit 0 embedded.

Before going any further, we would like to discuss why the 4-bit pattern is used to represent one watermarking bit. Actually, any n-bit pattern where n is greater than 3 can be used. In fact, a 3-bit pattern can be used as well, but, according to our experiments, if the 3-bit pattern is used, we cannot distinguish between the frame that is embedded the watermark bit 0 and the frame that is unembedded, as illustrated in Fig. 3.37. If we assume that there is no such an unembedded frame, then the 3-bit pattern works well. In this case, we can say one frame is divided into 3 subframes. If we want to embed the watermark bit 1, then "010" are embedded into the 3 subframes. If we want to embed the watermark bit 0, then "000" are embedded into those subframes. The reason that we have the first and the last zeros in the n-bit pattern is to ensure that all concavities are far enough. Therefore, without the assumption that there is no unembedded subframe, 4 is the minimum number of bits that serves the purpose.

To detect a watermarked frame, we define another scanner, which operates on $4M$ samples, called the frame-scan operation.

Given a watermarked audio signal $Y = [y_0\ y_1\ ...\ y_{H-1}]^{\mathrm{T}}$ of length H greater than $4M$, the frame-scan operation $\mathrm{F}[Y]$ scans from y_0 with a scan step of Δ until it detects the first watermarked frame.

Let $Y_i = [y_{i\Delta}\ y_{i\Delta+1}\ ...\ y_{i\Delta+4M-1}]^{\mathrm{T}}$ be a frame being scanned at step i, we first perform $\mathrm{S}[Y_i] = (b_0, b_1, ..., b_{\lfloor\frac{3M}{\delta}\rfloor})$. Then, four rectangular windows $W_j = (w_0^j, w_2^j, ..., w_{\lfloor\frac{3M}{\delta}\rfloor}^j)$ are separately perform on $\mathrm{Scan}[Y_i]$, i.e., $W_j * \mathrm{Scan}[Y_i] = (b_0^j, b_1^j, ..., b_{\lfloor\frac{3M}{\delta}\rfloor}^j)$ for $j = 1$ to 4.

$$w_k^j = \begin{cases} 1 & \text{if } k \in [\sigma, \sigma+\Sigma], \text{ for } j = 1, 2, 3, \\ 1 & \text{if } k \in [\sigma-\Sigma, \sigma], \text{ for } j = 2, 3, 4, \\ 0 & \text{otherwise,} \end{cases} \tag{3.15}$$

where $\sigma = \lfloor\frac{3M}{4\delta}\rfloor(j-1)$, and Σ is a positive integer, called the overlap margin.

The value of b_k^j is calculated by the following equation.

$$b_k^j = w_k^j \times b_k, \tag{3.16}$$

for $k = 0, 1, ..., \lfloor\frac{3M}{\delta}\rfloor$.

If $\sum_{k=0}^{\lfloor\frac{3M}{\delta}\rfloor} b_k^j$ is greater than $\lceil\frac{\Sigma}{2}\rceil$ for $j = 1$ or 4, or if $\sum_{k=0}^{\lfloor\frac{3M}{\delta}\rfloor} b_k^j$ is greater than $\Sigma+1$ for $j = 2$ or 3, we say that, by looking through the window W_j, the concavity of the singular spectrum is detected.

The scanner $\mathrm{F}[Y]$ will stop scanning and declare a watermarked frame only when the conditions described in Table 3.1 are satisfied. That is it will stop and return the extracted watermark bit 0 if and only if the concavity of singular spectrum can not be detected through the windows W_1, W_3 and W_4, but it can be detected through window W_2. It will return the extracted watermark bit 1 if and only if the concavity of the singular spectrum can not be detected through the windows W_1 and W_4, but it can be detected through window W_2 and W_3. Otherwise, it continues scanning with the step of Δ. Then, we can restart the frame-scan operation again and again until it reaches the end of the watermarked signal Y.

An example of performing the four windows to a frame $\mathrm{S}[Y_i]$ is shown in Fig. 3.38. The second and third frames are embedded so that the frame-scan operation can decode the pattern of b_i as the watermarked bit 1.

Table 3.1: Conditions for stopping frame-scan operation. Note that '$\circ$' is "The concavity of singular spectrum can be detected," and '$\times$' is "The concavity of singular spectrum cannot be detected."

	W_1	W_2	W_3	W_4
Watermark bit = 0	$\times$	$\circ$	$\times$	$\times$
Watermark bit = 1	$\times$	$\circ$	$\circ$	$\times$

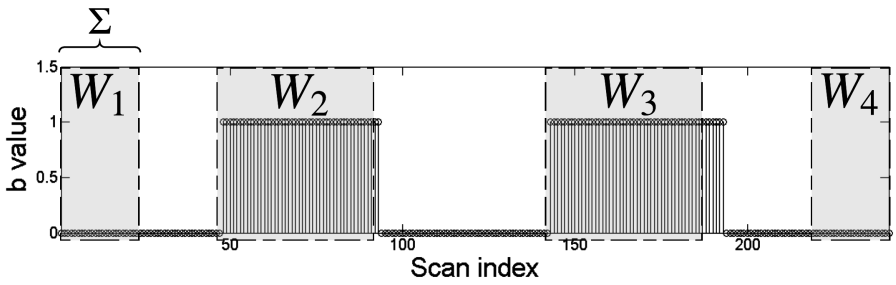

Figure 3.38: Example of performing the four windows to Scan$[Y_i]$.

3.6 AIH based on SSA in transform domain

As reviewed in Sect. 2.1.4 about the SVD-based audio watermarking schemes, some techniques form matrices in certain transform domains, such as wavelet-, discrete cosine-, and Fourier-transform domains. Even though the published results from both time and transform domains are comparable [19, 23, 36], it is worthwhile to investigate the SSA in other domain. In addition, to our knowledge, there is no AIH based on SSA in transform domains in existing literature.

In this section, we study and construct the AIH scheme based on the SSA in discrete wavelet transform (DWT) domain. We made an assumption in the beginning that, since the human auditory system is not sensitive to high frequency [129], it should be possible to modify the very-low-order singular values of the trajectory matrix that is constructed by the detail coefficients of the DWT. As discussed in Sects. 2.2.2 and 3.2.2, the lower the order, the better performance in robustness.

3.6.1 Discrete wavelet transform

A wavelet transform is a tool that cuts up signals into different frequency components, and then studies each component with a resolution matched to its scale [130]. That is, given a signal of finite energy, the wavelet transform projects the signal on a family of frequency bands which are generated by the shifts of the mother wavelet. A DWT is any wavelet transform where the wavelet are discretely sampled [131]. The advantages of the wavelet transform are as follows [132]: (1) it offers a simultaneous localization in time and frequency domains, (2) it is computationally very fast with the fast wavelet transform, (3) it is able to separate the fine details in a signal, (4) it can be used to decompose a signal into component wavelet, and (5) it can obtain a good approximation of the given function by using only a few coefficients.

The DWT of a signal x is calculated by passing the signal x through cascading filter bank. An example of 3-level DWT is shown in Fig. 3.39. The signal $x[n]$ is decomposed simultaneously by using a low-pass filter with impulse response g and a high-pass filter with impulse response h. Then, half the samples from each filter can be deleted because half the frequencies of the signal $x[n]$ are removed. The outputs from the low-pass filter branch is called approximation coefficients, whereas those from the high-pass filter branch is called detail coefficients. The approximation coefficients from the preceding level are then passed through the same pair of filter bank at the next level, and so on. The frequency domain representation of the three-level DWT of the signal with N samples, frequency range from 0 to f_n is illustrated in Fig. 3.40. According to this example, four output scales are produced: (1) detail coefficients of $\frac{N}{2}$ samples with frequency range from $\frac{f_n}{2}$ to f_n from the first level, (2) detail coefficients of $\frac{N}{4}$ samples with frequency range from $\frac{f_n}{4}$ to $\frac{f_n}{2}$ from the second level, (3) detail coefficients of $\frac{N}{8}$ samples with frequency range from $\frac{f_n}{8}$ to $\frac{f_n}{4}$ from the third level, and (4) approximation coefficients of $\frac{N}{8}$ samples with frequency range from 0 to $\frac{f_n}{8}$.

3.6.2 Embedding process

The embedding process consists of eight steps, as shown in Fig. 3.41 (left).

1. The host signal is segmented into non-overlapping frames.

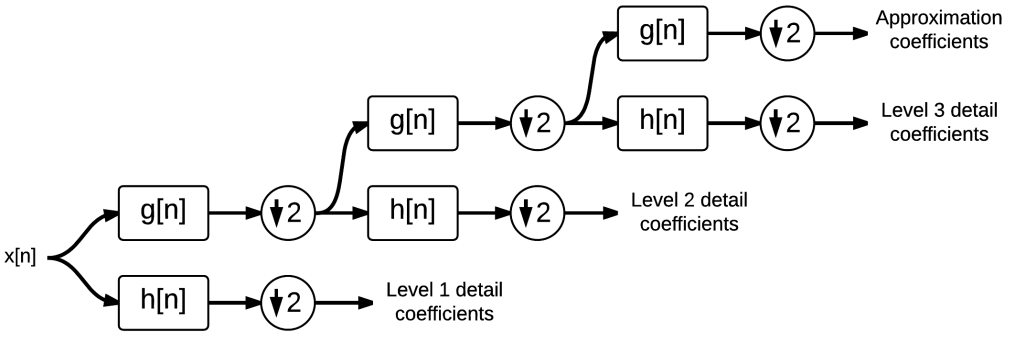

Figure 3.39: Three-level DWT filter bank.

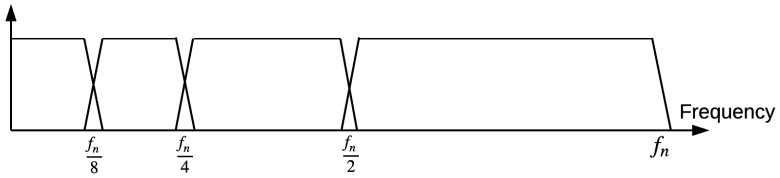

Figure 3.40: Frequency domain representation of the three-level DWT of the signal with frequency range from 0 to f_n.

2. Each frame is represented by the approximation and detail coefficients by using the DWT.
3. Only detail coefficients are used to construct the trajectory matrix.
4. SVD is performed to each matrix.
5. The singular values obtained from the previous step with the index within the interval of $[u+1, l-1]$ are forced to have the same value as the singular value at the index u if the watermark bit is 1; otherwise, their new values are the same as the singular value at the index l.
6. The modified detail coefficients are created by Hankelizing the modified trajectory matrix.
7. The watermarked frame is built by applying inverse DWT to the approximation and modified detail coefficients.
8. Finally, the watermarked signal is obtained by stacking all frames.

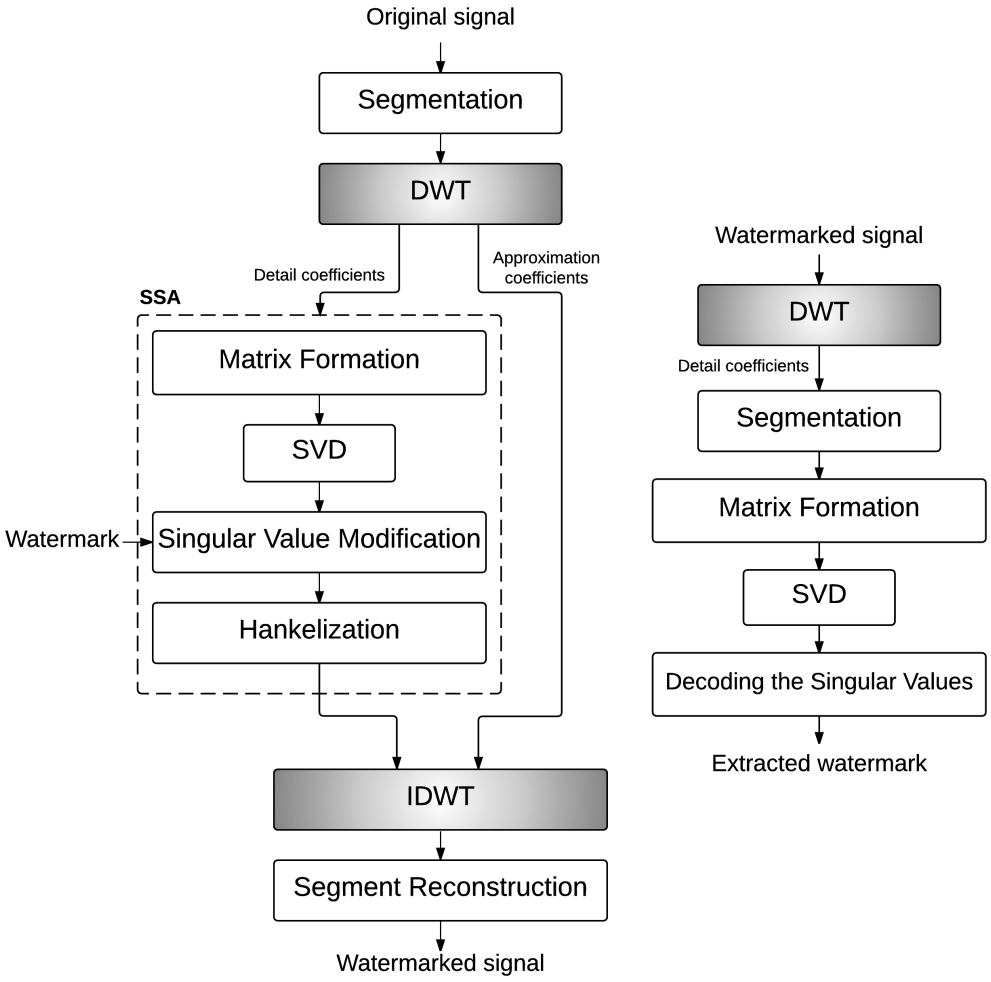

Figure 3.41: Embedding and extraction processes of the AIH based on SSA in DWT domain.

3.6.3 Extraction process

The extraction process consists of five steps, as shown in Fig.3.41 (right). The first four steps are exactly the same as those of the embedding process. After obtaining the singular values, the method of decoding by the median, as described in Sect. 3.2.3, is adopted in the final step.

3.7 Summary

This chapter proposed the SSA-based AIH frameworks. The basis of all frameworks is called the core structure. The embedding process of the core structure modifies some predefined singular values. These modified singular values are signal-independent. We also discussed the effect of embedding areas on the sound quality of watermarked signal and the robustness of a watermark. To discuss the effect on the sound quality, we defined the strain as a measure of deformation of a singular value under attacks. We showed that a higher-order singular value has a higher strain, thus, it is less robust, but the sound quality of a watermark signal is better.

Then, we proposed the improved framework in which the differential evolution optimizer is integrated. The optimizer is used to find the parameters. Since the modified singular values are signal-dependent, the automatic parameter estimation was proposed to find those parameters in the extraction process; otherwise, the framework cannot be considered as blind.

Then, we proposed the other improved framework in which the psychoacoustic model 1 is integrated. The relationship between singular-value indices and frequency components was established.

Then, we proposed the singular-spectrum analyzing technique that can be used to automatically detect watermarked frames. Finally, we studied the basic SSA in DWT domain.

Chapter 4

Implementations and evaluations of the proposed schemes

Based on the frameworks described in the previous chapter, we implemented and evaluated three related, but different, models. We start with a database, conditions, and evaluation methods in the first section, after that the parameters that were used to implement each model are given in detail. The automatic frame detection and the AIH scheme based on SSA in DWT domain were also implemented and investigated their properties in some aspects. Finally, we demonstrate the AIH scheme in which a psychoacoustic model is incorporated.

4.1 Database and evaluation methods

4.1.1 Database and conditions

One hundred host signals from the RWC music-genre database [133] were used in our experiments. All have a sampling rate of 44.1 kHz, 16-bit quantization, and two channels. A watermark was embedded in one channel. The audio-signal duration depends upon the capacity, total number of watermark bits, and the number of repetitions. Unless otherwise stated, the watermark was embedded starting from the initial segment of host signals. All simulations were operated on the Fujitsu CX250 Cluster (JAIST parallel computers), where each computing node is equipped with two Intel Xeon E5-2680v2 2.80 GHz (10 cores each) and 64 GB memory (4 GB DDR3-1866 ECC $\times$16).

We compared the proposed schemes with the conventional SVD-based scheme [23]. It was chosen as a reference for three reasons:

1. It is one of a few blind SVD-based techniques.
2. Its published results are promising.
3. Both the SSA-based and SVD-based frameworks belong to the same class of the audio information hiding.

4.1.2 Sound-quality tests

The purpose of sound-quality tests is to measure the effectiveness of the proposed schemes in terms of inaudibility. Both objective and subjective evaluations were conducted in our experiments.

Objective evaluation

Three following distance measures the perceptual evaluation of audio quality (PEAQ), the LSD, and the SDR were chosen. The PEAQ measures the degradation of the watermarked signal compared with the original. It covers a scale, so called the *objective difference grade* (ODG) from -4 (worst) to 0 (best) [134]. The LSD is a distance measure between two spectra and is defined in Eq. (3.5). The SDR is a power ratio between the signal and the distortion and is defined in Eq. (3.12).

The evaluation criteria for the good sound-quality are as follows. The ODG must be greater than -1 (not annoying), the LSD must be less than 0.4 dB, and the SDR must be greater than 20 dB.

Before evaluating the proposed schemes, we have already verified the perfection of SSA as an analysis-synthesis tool by checking the ODG, LSD, and SDR of synthesis signals (without embedding), compared to the originals. The results showed that all ODGs were around zero, all LSDs were zero, and all SDRs were infinity [135]. In other words, the framework of singular-spectrum analysis is perfect in the sense that the framework itself does not introduce distortion into the system.

Subjective evaluation

The ABX test was used to evaluate sound quality of watermarked signals. The ABX test was chosen because it is a discrimination test method for the overall difference when there is no specified attribute to be discriminated [136]. The ABX test is to identify whether a listener can perceive a difference between two clips of music or not. Firstly, we presented a subject with two clips, A and B, and then, after presenting the third clip X, which is randomly selected from either A or B, we asked the subject to match X to either A or B. In our experiments, A was the original signal, and B was the watermarked signal. The sequence of presenting A and B to participants was random. We tested 20 audio clips in total. There were 35 normal-hearing subjects participating in the test. We assumed a binomial distribution and used it as a statistical analysis tool [137]. Typically, the 95% confidence level is sufficient for psychoacoustic experiments, i.e., to reach such a level, 14 out of 20 correct identifications was the criterion indicating a perceptible difference [138].

4.1.3 Robustness tests

The effectiveness of the proposed schemes in terms of robustness is measured by the watermark extraction precision. Either the bit-error rate (BER) or bit-detection rate (BDR) can represent the watermark extraction precision. The BER is defined as the number of error bits divided by the total number of embedded bits. Given the embedded-watermark bit-string $w(i)$ and the extracted-watermark bit-string $\hat{w}(i)$ for $i=1$ to N,

$$\mathrm{BER} = \frac{\sum_{i=1}^{N} w(i) \oplus \hat{w}(i)}{N}, \tag{4.1}$$

where $\oplus$ is the bitwise XOR operator.

In this work, the BER is preferred, and the relation between the BER and BDR is that $\mathrm{BER} = 1 - \mathrm{BDR}$. The criterion for the robust AIH is that the BER must be lower than 0.1 or 10%.

Five attacks were performed on watermarked signals: (1) Gaussian-noise addition with average signal-to-noise ratio (SNR) of 36 dB, (2) re-sampling with 16 and 22.05 kHz, (3) band-pass filtering with 100-6000 Hz and -12 dB/Oct, (4) MP3 compression with 128 kbps joint stereo, and (5) MP4 compression with 96 kbps.

Table 4.1: Parameters for the fixed-parameter model.

	Simulation 1	Simulation 2	Simulation 3
Frame/subframe length N (sample)	2450	816	816
Window length L (sample)	980	326	326
Parameter u	20	20	20
Parameter l	60	60	60
Parameter ϵ	0.1	0.1	0.1
Embedding capacity (bps)	18	54	10.8
Payload (bit)	150	150	150
Total duration (second)	8.3	2.8	13.9
Repetition	No	No	5

Table 4.2: ODGs, LSDs, and SERs: comparison of the fixed-parameter model and the conventional method.

	ODG	LSD	SDR
Simulation 1	−0.13	0.36	20.96
Simulation 2	−0.35	0.68	23.49
Simulation 3	−0.13	0.63	23.01
Conventional method [23]	0.20	0.11	26.82

4.2 Implementations and evaluations of the SSA-based AIH schemes

4.2.1 Fixed-parameter model

The fixed-parameter model was implemented based on the core structure of the SSA-based AIH schemes, as described in Sect. 3.2, with the parameters shown in Table 4.1.

Sound-quality evaluation

The average ODG, LSD, and SDR of watermark signals compared between the fixed-parameter model and the conventional SVD-based method are shown in Table 4.2. Note that the capacity of the conventional SVD-based method is around 6 bps, and its frame size is 85×85 samples. The evaluation details can be found in Appendix A.

Table 4.3: BERs (%) comparison of the fix-parameter model and the conventional SVD-based method when attacks (i.e., MP3 and MP4 compression, Gaussian noise addition (AWGN), re-sampling with 16 and 22.05 kHz (RES 16 and RES 22.05, respectively), and band-pass filtering (BPF)) were performed.

	No attack		MP3		MP4		AWGN		RES 16		RES 22.05		BPF	
	POL	MED	POL	MED	POL	MED	POL	MED	POL	MED	POL	MED	POL	MED
Simulation 1	1.76	1.17	7.51	7.10	5.03	4.35	1.76	1.17	4.17	3.78	2.76	2.26	4.95	5.91
Simulation 2	4.10	2.24	16.38	14.77	11.52	9.28	4.10	2.24	7.54	6.14	5.11	3.54	12.92	11.28
Simulation 3	0.97	0.13	14.59	11.63	9.61	6.74	0.98	0.17	5.22	4.24	2.48	1.78	10.52	8.25
Conventional method [23]	0.00		59.00		1.20		0.00		2.67		2.67		38.08	

Table 4.4: Average BERs (%): comparison of the fix-parameter model and the conventional SVD-based method.

	Average BER (%)	
	POL	MED
Simulation 1	3.99	3.68
Simulation 2	8.81	7.07
Simulation 3	6.33	4.70
Conventional method [23]	14.80	

Robustness evaluation

The results of robustness tests are shown in Table 4.3. We also compare the BERs when the watermark is decoded by the median method (MED) and by the polynomial fitting method (POL). The average BERs of the fixed-parameter model and the conventional method are summarized in Table 4.4. The evaluation details can be found in Appendix A.

Discussion

In this work, we compare our results with the conventional SVD-based method [23]. The details of comparative evaluations of the conventional method, the other typical methods, and the fixed-parameter model can be found in Appendix D. In short, the conventional SVD-based method is used as our benchmark because it is the most robust method against several attacks, except against the MP3 compression and the band-pass filtering. In addition, it is blind and inaudible.

Compared with the conventional method, the degradation in the sound quality of

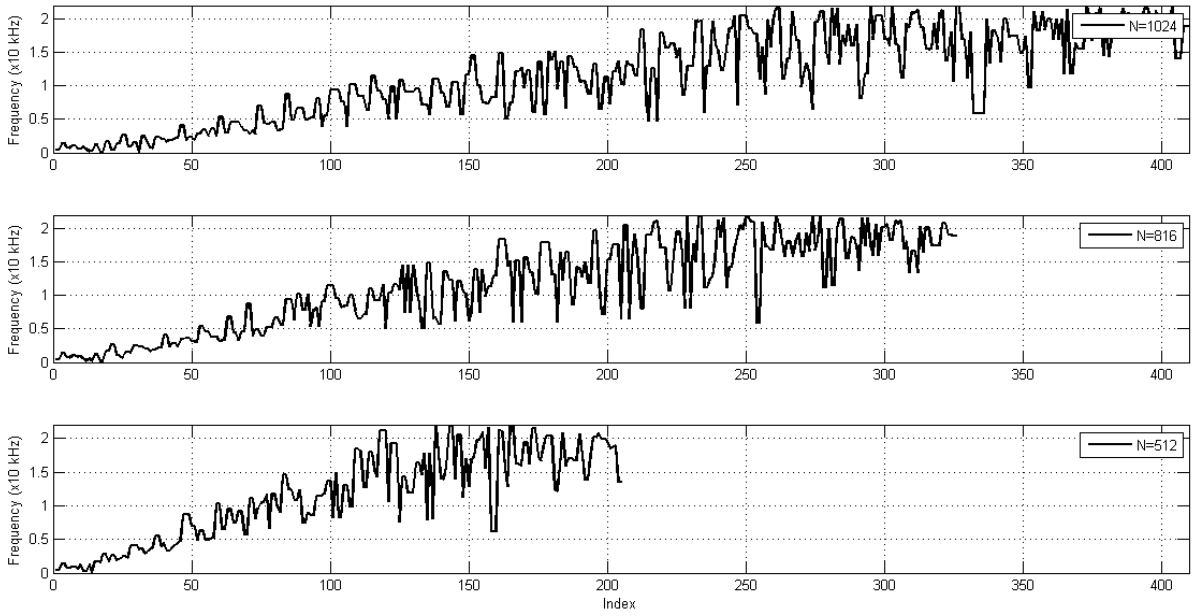

Figure 4.1: Relation between frequencies and singular-value indices at different frame size.

watermarked signals obtained from the fixed-parameter model is quite larger, especially when the frame or subframe size of the SSA-based method is small, as in the cases of simulations 2 and 3. Effect of the frame size to sound quality is not beyond our expectations because the smaller frame size implies the higher capacity. This relationship is supported by comparing the results from the simulations 1 and 2, where the capacity and the ODG and LSD are inversely proportional. There is another subtle reason that the frame size affects sound quality. It associates with our findings concerning the relation between frequency ranges and singular-value indices. When the frame size is smaller, the singular-value interval covers the larger range of frequencies, as shown in Fig. 4.1. Therefore, modifying singular values of the smaller frame size affects more frequency components than modifying those of the larger frame size.

On the other hand, the robustness of the fixed-parameter model is better than that of the conventional one. However, we have been aware that this conclusion might be biased due to the possibility that the conventional method might be fragile only to the MP3 and band-pass filtering since the number of attacks is not big enough. However, the BERs of the simulation 1 and the conventional method are comparable even when we do not take the MP3 and band-pass filtering into consideration. We also found that there is no significant difference in terms of extraction performance between the median and

polynomial fitting methods for the fixed-parameter model.

Comparing the BERs of simulation 2 and 3 can confirm that BER can be reduced by embedding repetitions, as theoretically explained in Sect. 3.2.4. Compared the frame sizes in the simulation 1 and 3, we can infer that embedding information in one long frame is better than doing so in many short frames of the same total duration in terms of both inaudibility and robustness. However, when computation complexity is taken into account, embedding repetitions may be better because time complexity of SVD of a $m \times n$ matrix is $O(\min\{mn^2, m^2n\})$ [139–141]. That is, embedding x subframes of length N is less complex than embedding a frame of length $x \times N$. For this very reason, in our latter implementations, even though we could make a better result by increasing the frame size, we chose to implement based on comparatively small frames, and it is enough to verify the concepts of SSA-based AIH.

The final remark is about the robustness of the host signal in the sense that whether all audio signals have the same tendency, e.g., the properties of the singular values, in the SSA-based decomposition or not. According to the primary theorem concerning the SVD, every matrix of complex numbers has a singular value decomposition, and the singular values are uniquely determined [142]. Therefore, this theorem guarantees the existence of the singular values. The only property of the singular-value plot that we require is that the plot is approximately convex. Mathematically, the region above the singular spectrum plot is not a convex set. However, we do not require the plot to be strictly convex. Broadly speaking, we require that the approximation of the singular spectrum curve by the quadratic equation is convex. From our experiments, and based on the robustness evaluation results, this condition holds. It is possible that the non-strictly-convex nature of the singular spectrum plot might pose a problem that limits the overall performance of the scheme; therefore, this issue will be investigated further.

4.2.2 Partially-blind, adaptive-parameter model

The partially-blind, adaptive-parameter model (hereinafter called the *partially-blind model*) was implemented based on the core structure where the parameter set $\{u, l, \epsilon\}$ was determined by the differential evolution optimization, as described in Sect. 3.3.1. The model is partially-blind because it is assumed that the parameters, which are signal-dependent, are

Table 4.5: Parameters for the partially-blind model.

Subframe length N (sample)	2450
Window length L (sample)	980
Embedding capacity (bps)	18
Payload (bit)	150
Total duration (second)	8.3
Repetition	No
$\{u, l, \epsilon\}$	provided in Appendix B

Table 4.6: Average, maximum, and minimum correct identification of the ABX tasks with 20 stimuli for the fixed-parameter and partially-blind models.

	Fixed-parameter model	Adaptive-parameter model
Average correct identification (%)	73.00	49.78
Maximum correct identification	19	13
Minimum correct identification	9	7

presented at the extraction stage. In other words, the parameters u, l, and ϵ are shared between the embedding and extraction processes, someway. The aim of this implementation is to verify the assumption that the poor sound-quality of some watermarked signals obtained from the fixed-parameter model is due to the singular-value modification that takes place in inappropriate locations. The parameters for the implementation of the partially-blind model are shown in Table 4.5.

Sound-quality evaluation

The results from objective sound quality tests are shown in Table 4.7. Twenty out of 100 audio clips were randomly selected and used in the ABX test with 35 normal-hearing persons. The results are shown in Table 4.6. The correct identification of X when audio signals were embedded by using the fixed-parameter model was around 70%, whereas it was about the chance level in the case of the adaptive-parameter model.

Robustness evaluation

The results from robustness tests are shown in Table 4.8. The table shows only BERs from the median method. The evaluation details can be found in Appendix A.

Table 4.7: ODGs, LSDs, and SERs: comparison of the fixed-parameter model, partially-blind model, and the conventional method.

	ODG	LSD	SDR
Partially-bland model	0.19	0.16	35.25
Fixed-parameter model	−0.13	0.36	20.96
Conventional method [23]	0.20	0.11	26.82

Table 4.8: BERs (%): comparison of the fix-parameter model, partially-blind model, and the conventional SVD-based method when attacks (i.e., MP3 and MP4 compression, Gaussian noise addition (AWGN), re-sampling with 16 and 22.05 kHz (RES 16 and RES 22.05, respectively), and band-pass filtering (BPF)) were performed.

	No attack	MP3	MP4	AWGN	RES 16	RES 22.05	BPF
Partially-blind model	1.34	5.80	6.52	1.34	2.58	1.92	5.62
Fixed-parameter model	0.13	11.63	6.74	0.17	4.24	1.78	8.25
Conventional method [23]	0.00	59.00	1.20	0.00	2.67	2.67	38.08

Discussion

Compared with the fixed-parameter model, the robustness of the partially-blind model is comparable. However, the sound quality significantly improves both in objective and subjective evaluations. The average correct identification of 49.78% indicates that subjects hardly perceive the difference between A and B, i.e., between the original and watermarked signals.

From our experiments, we found that the parameter set obtained from differential evolution was good at some specific tasks included in the differential evaluation optimization. For example, if the simulation of MP3 compression is removed from the optimizer in Fig.3.14, watermarked signals will be fragile to the MP3 attack. Therefore, the practical model should include potential attacks as many as possible in order to find the best parameters. Note that adding an attack to the optimizer increases computational time considerably, so it presents some kind of trade-off.

Table 4.9: Parameters for the completely-blind model.

Subframe length N (sample)	2450
Window length L	980
Embedding capacity (bps)	18
Payload (bit)	150
Total duration (second)	8.3
Repetition	No
$\{u, l, \epsilon\}$	provided in Appendix B

Table 4.10: ODGs, LSDs, and SERs: comparison of the fixed-parameter model, partially-blind model, completely-blind model, and the conventional method.

	ODG	LSD	SDR
Completely-blind model	0.18	0.24	22.39
Partially-blind model	0.19	0.16	35.25
Fixed-parameter model	−0.13	0.36	20.96
Conventional method [23]	0.20	0.11	26.82

4.2.3 Completely-blind, adaptive-parameter model

Similar to the partially-blind model, the completely-blind, adaptive-parameter model (hereinafter called the *completely-blind model*) was built based on the core structure and the differential evolution. In addition, to avoid the assumption that the extraction process knows the signal-dependent parameters in advance, the automatic parameter estimation described in Sect. 3.3.2 was integrated into the scheme.

From our preliminary comparative-evaluation of the derivative-based and concavity density-based methods, we found that the latter provides a better estimation (see Appendix C). Hence, it was chosen to estimate the parameter set $\{u, l, \epsilon\}$ in the extraction process. The parameters for the implementation of the completely-blind model are shown in Table 4.9.

Sound-quality evaluation

The results from sound-quality tests are shown in Table 4.10. The evaluation details can be found in Appendix A.

Table 4.11: BERs (%): comparison of the fix-parameter model, partially-blind model, completely-blind model, and the conventional SVD-based method when attacks (i.e., MP3 and MP4 compression, Gaussian noise addition (AWGN), re-sampling with 16 and 22.05 kHz (RES 16 and RES 22.05, respectively), and band-pass filtering (BPF)) were performed.

	No attack	MP3	MP4	AWGN	RES 16	RES 22.05	BPF
Completely-blind model	11.45	21.54	13.74	11.45	15.64	13.76	22.77
Partially-blind model	1.34	5.8	6.52	1.34	2.58	1.92	5.62
Fixed-parameter model	0.13	11.63	6.77	0.17	4.24	1.78	8.25
Conventional method [23]	0.00	59.00	1.20	0.00	2.67	2.67	38.08

Robustness evaluation

The results of robustness tests are shown in Table 4.11. The evaluation details can be found in Appendix A.

Discussion

Compared with the partially-blind model, the robustness of the completely-blind model decreases. On the other hand, compared with the fixed-parameter model, the sound quality of the completely-blind model increases. Its robustness is poorer than that of the partially-blind model because there is one additional and critical condition that constrains the differential evolution, i.e., the extraction process must be blind. From our experimental results, to be blind, the intervals $[u, l]$ have to be large. As a result, a large interval is not likely to be placed at low-order singular values since it will strongly affect the sound quality. For this reason, most parameters u and l are quite large compared with those delivered by the differential evolution in the partially-blind model. Therefore, it is less robust. The completely-blind model is better than the fixed-parameter in sound quality for a very obvious reason. It deploys the differential evolution.

4.3 Experiments on AIH integrated with a psychoacoustic model

The following schemes were implemented based on the core structure, the psychoacoustic model 1, and the frequency-to-index mapping algorithm described in Sect. 3.4. The objective of these implementations was to verify that the psychoacoustic model can provide the SMR information that the SSA-based AIH scheme can use to improve the performance. In this section, we report the results from two implementations. The first one is the implementation without the automatic parameter estimation, which can be considered as the partially-blind scheme. The second one is the implementation with the automatic parameter estimation, which can be considered as the completely-blind scheme.

4.3.1 Scheme without the automatic parameter estimation

We simulated the scheme without the automatic parameter estimation in four conditions on the SMR criterion γ and the frame length N. The parameters for this implementation are shown in Table 4.12.

We tested with 12 signals randomly selected from the RWC database. The comparison of the average ODGs, the average LSDs, and the average SDRs are shown in Table 4.13. It can be seen from this table that there is no significant difference in sound quality among the four simulations, and they are slightly poorer than the partially-blind model and the conventional SVD-based one in the LSD. However, in other measures (the ODG and the SDR), they are comparable or better than them. Therefore, in terms of inaudibility, the psychoacoustic model can be used to determine the parameters as well as the differential evolution.

The results from robustness tests are shown in Table 4.14. The average BER of the first, second, third, and the fourth simulations, the fixed-parameter model, the partially-blind model, and the conventional SVD-based method are 30.78%, 9.56%, 5.03%, 4.76%, 3.54%, 3.58%, and 14.80%, respectively. Except for the first simulation, the other proposed schemes are better than the conventional one. Comparing all four simulations, we found that the smaller the frame size, the poorer the robustness. The robustness depends upon the value of γ. The reason for this result is that the low SMRs imply inaudible

Table 4.12: Parameters for the psychoacoustic-model-based AIH schemes. Note that the word *adaptive* means the value of γ depends on the maximum SMR. If the maximum SMR is greater than 25 dB, $\gamma = 18$. If the maximum SMR is less than 20 dB, $\gamma = 12$. Otherwise, $\gamma = 15$. N is the audio frame-length into which one bit of the hidden information is embedded. Also, note that this frame size is not necessary to be the same as that of the psychoacoustic model.

	Simulation 1	Simulation 2	Simulation 3	Simulation 4
Frame length N (sample)	816	2450	2450	2450
Window length L (sample)	326	980	980	980
Embedding capacity (bps)	54	18	18	18
Payload (bit)	150	150	150	150
Total duration (second)	2.77	8.33	8.33	8.33
Repetition	No	No	No	No
SMR criterion γ (dB)	10	10	15	*adaptive*
Parameters u and l	provided in Appendix B			

components. The perceptual-coding techniques, such as the MP3 or MP4 compression, normally destroy components with the low SMRs. Thus, embedding hidden information into the components with the higher SMRs increases the chance that the information survives after signal processing. Another reason is that the higher SMRs usually associate with the lower singular-value indices. We also found that embedding the information in the lower index is more robust. When the adaptive strategy is used to set the value of γ, its average BER is very close to those of the previous SSA-based methods. Therefore, with a suitable value of γ, the psychoacoustic model can be used to deliver the parameters as well as the differential evolution in terms of robustness.

4.3.2 Scheme with the automatic parameter estimation

The automatic parameter estimation, as described in Sect. 3.3.2, was integrated into the AIH scheme with the psychoacoustic model. In this experiment, we used the same settings as those of the simulation 4 in Table 4.12. Since the automatic parameter estimation does not affect the embedding process of the scheme, the sound-quality results were the same as those of the simulation 4 in Table 4.13. However, the robustness results were different because they depended on the accuracy of the automatic parameter estimation. The

Table 4.13: ODGs, LSDs, and SDRs: comparison of the psychoacoustic-model-based AIH schemes, the fixed-parameter model, the partially-blind model, and the conventional SVD-based method.

	ODG	LSD	SDR
Simulation 1	0.20	0.33	42.52
Simulation 2	0.19	0.22	34.75
Simulation 3	0.16	0.31	25.68
Simulation 4	0.18	0.34	24.30
Fixed-parameter model	-0.13	0.36	20.96
Partially-blind model	0.19	0.16	35.25
Conventional method [23]	0.20	0.11	26.82

Table 4.14: BERs (%) comparison of the psychoacoustic-model-based AIH schemes, the fixed-parameter model, the partially-blind model, and the conventional SVD-based method when attacks (i.e., MP3 and MP4 compression, Gaussian noise addition (AWGN), re-sampling with 16 and 22.05 kHz (RES 16 and RES 22.05, respectively), and band-pass filtering (BPF)) were performed.

	No attack	MP3	MP4	AWGN	RES 16	RES 22.05	BPF
Simulation 1	4.17	31.33	47.94	4.17	52.39	22.11	53.33
Simulation 2	2.00	22.50	12.83	2.00	9.50	4.39	13.72
Simulation 3	1.67	11.44	5.39	1.67	5.89	2.50	6.67
Simulation 4	1.61	10.94	5.22	1.61	5.17	2.39	6.39
Fixed-parameter model	1.17	7.10	4.35	1.17	3.78	2.26	4.95
Partially-blind model	1.34	5.80	6.52	1.34	2.58	1.92	5.62
Conventional method [23]	0.00	59.00	1.20	0.00	3.67	1.67	38.08

Table 4.15: Actual and estimated values of the parameters u and l used in the scheme with the automatic parameter estimation. The most left column shows the track numbers.

	Actual value		Estimated value	
	u	l	$\hat{u}$	$\hat{l}$
Track no. 01	30	90	26	90
Track no. 07	40	90	39	93
Track no. 13	20	60	23	66
Track no. 28	45	110	43	110
Track no. 37	40	120	44	120
Track no. 49	40	100	38	99
Track no. 54	30	90	30	90
Track no. 57	75	160	81	155
Track no. 64	30	90	28	90
Track no. 85	60	140	64	140
Track no. 91	40	100	37	100
Track no. 100	35	93	38	95

estimated values of u and l, compared with the actual ones, are shown in Table 4.15.

The results from the robustness evaluations are shown in Table 4.16. The average BERs of the psychoacoustic-model-based scheme without and with the automatic parameter estimation, the fixed-parameter model, the partially-blind model, and the conventional SVD-based method are 4.76%, 6.78%, 3.54%, 3.58%, and 14.80%, respectively. The average BER of the scheme with the automatic parameter estimation satisfies the criterion for the robustness as its average BER was less than 10%. Compared with the conventional method, the psychoacoustic-model-based schemes are more robust. However, compared with the previously proposed SSA-based methods, they are less robust to some degree. Therefore, the overall performance of the psychoacoustic-model-based schemes seem to be slightly poorer than that of the partially-blind model.

Table 4.16: BERs (%) comparison of the psychoacoustic-model-based AIH schemes with and without the automatic parameter estimation (APE), the fixed-parameter model, the partially-blind model, and the conventional SVD-based method when attacks (i.e., MP3 and MP4 compression, Gaussian noise addition (AWGN), re-sampling with 16 and 22.05 kHz (RES 16 and RES 22.05, respectively), and band-pass filtering (BPF)) were performed.

	No attack	MP3	MP4	AWGN	RES 16	RES 22.05	BPF
Scheme with the APE	2.50	19.44	6.61	2.50	8.83	6.06	11.50
Scheme without the APE	1.61	10.94	5.22	1.61	5.17	2.39	6.39
Fixed-parameter model	1.17	7.10	4.35	1.17	3.78	2.26	4.95
Partially-blind model	1.34	5.80	6.52	1.34	2.58	1.92	5.62
Conventional method [23]	0.00	59.00	1.20	0.00	3.67	1.67	38.08

4.3.3 Discussion

There are five observations concerning the performance and the limitation of the proposed AIH schemes with the psychoacoustic model to be discussed. First, we have shown that the psychoacoustic model can be used to determine the parameters u and l. These parameters are host-signal-dependent and of importance because their values determine the balance between the inaudibility and the robustness. In the partially-blind model, the differential evolution is used to determine the parameters. Compared with using the differential evolution, there are three great advantages of using the psychoacoustic model.

- The computational time is reduced considerably because the differential evolution optimization has a large search space. The comparison of the computational time is shown in Table 4.17. To determine the parameters u and l for one signal, the differential evolution takes about 13 hours, whereas the psychoacoustic-model-based method takes about 4.3 seconds.
- The optimal parameters from the differential evolution depends on many factors, such as the simulations included in the optimizer. Moreover, the cost function has two additional parameters. In this sense, using the psychoacoustic model reduces the number of scheme parameters.
- We can achieve the better inaudibility when the frame size is small, as shown in Table 4.18. It implies that the psychoacoustic-model-based scheme can achieve a

Table 4.17: Comparison of the computational times for determining the parameters of a host signal when the automatic parameterization is based on the differential evolution and when it is based on the psychoacoustic model.

	Function	Computational time of the function	Search space/ No. of operation	Approximated total computational time
Differential Evolution	Cost function evaluation	3 minutes 44 seconds	31815 possible vectors	13 hours 9 minutes
Psychoacoustic model	SMR calculation	0.36 seconds	717 times	4.3 minutes

higher embedding capacity, compared with the partially-blind model.

Second, in the first experiment, the fourth simulation is the best, compared with other simulations. However, the robustness of this simulation is slightly poorer than that of the fixed-parameter model and the partially-blind model. This is because we use only the SMR from the psychoacoustic model as the guidance for determining the parameters. The low SMR can gain inaudibility but may lose robustness. In addition, the low-SMR component is more likely to be destroyed by the perceptual coding such as the MP3 compression. As a result, the BER of the fourth simulation is marginally higher than that of the previously proposed schemes when the MP3 compression is performed on the watermarked signals. To improve the robustness, we may include the two-tone suppression to the psychoacoustic model. Then, the hidden information is embedded into the suppressed areas where they have a high SMR so that the scheme can achieve both inaudibility and robustness. This is one of our future work.

Third, different from the previously proposed schemes, in this one, we also investigated the effect when we slightly adapted the embedding rule. Instead of modifying the singular spectrum when the watermark bit 0 is embedded, we did not modify it. We found that the effectiveness in terms of robustness are the same, but in terms of inaudibility, the objective scores improves slightly, as shown in Table 4.19. The previously proposed schemes, especially the one with the differential evolution optimization, can benefit from this fact because the optimization function directly handles the trade-off between inaudibility and robustness.

Fourth, analyzing a host signal by using the psychoacoustic model and determining the relationship between the frequency and the singular-value index of the signal might expose the limitation in finding the balance between the inaudibility and the robustness of

Table 4.18: ODGs, LSDs, and SDRs: comparision of the psychoacoustic-model-based AIH scheme and the partially-blind model when the frame size is small.

	ODG	LSD	SDR
Simulation 1	0.20	0.33	42.52
Partially-blind model ($N = 816$)	-0.35	0.68	23.49

Table 4.19: ODGs, LSDs, and SDRs: comparison of the psychoacoustic-model-based AIH scheme when the singular spectrum is modified and when it is not modified to embed the watermark bit 0.

	ODG	LSD	SDR
The singular spectrum is modified.	0.18	0.34	24.30
The singular spectrum is not modified.	0.18	0.25	25.61

the proposed scheme. Figures 4.2 and 4.3, for instance, show the psychoacoustic analysis of a host signal and the relationship between the peak frequency and the singular-value index of the host signal, respectively. Figure 4.2 shows that components with low SMRs are under the absolute threshold of hearing. On the one hand, if we embed information into these components, the information can be easily destroyed by the perceptual-coding techniques. Therefore, it is not robust. On the other hand, if we embed information into the components with higher SMRs, the inaudibility may not be achieved. The proposed psychoacoustic-model-based scheme encounters difficulty when the frame has this kind of characteristic. The similar difficulty can be seen from a different viewpoint, as shown in Fig. 4.3. In this figure, a small singular-value interval associates with a large frequency-range. When we modify a large interval on a singular spectrum to achieve a good robustness, it will strongly affect many frequency-components so that the sound quality may drop.

The final remark is about the relationship between the singular-value index and the peak frequency of the oscillatory components. This relationship is unique because the singular values are uniquely determined. That is, given one frame, we have only one curve that describes the relation between the singular-value index and the peak frequency of the oscillatory components associated with that index. However, these relationships vary from frame to frame. The examples of these relationships from six different frames

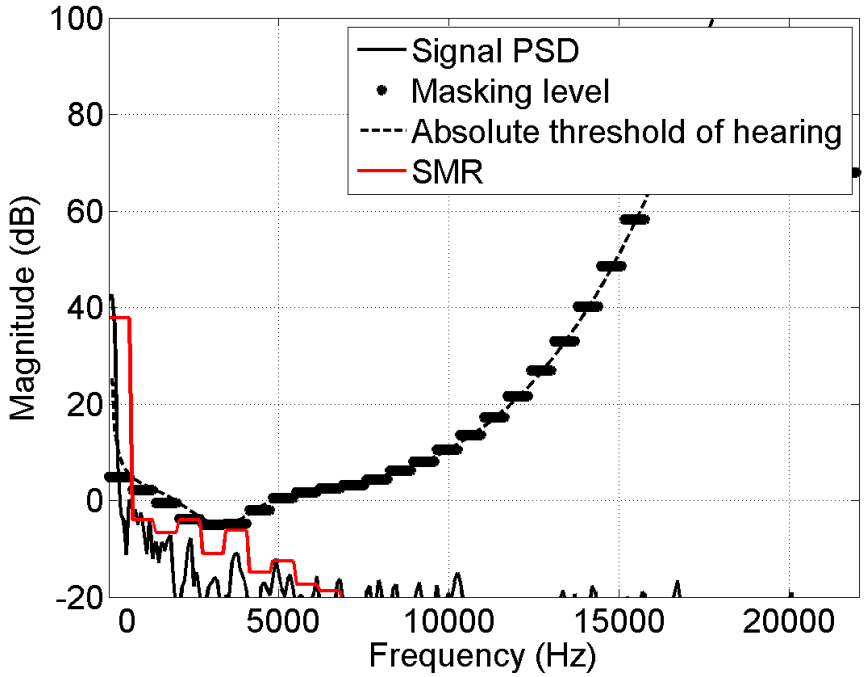

Figure 4.2: Psychoacoustic analysis of the host signal (track no. 57) and its SMR.

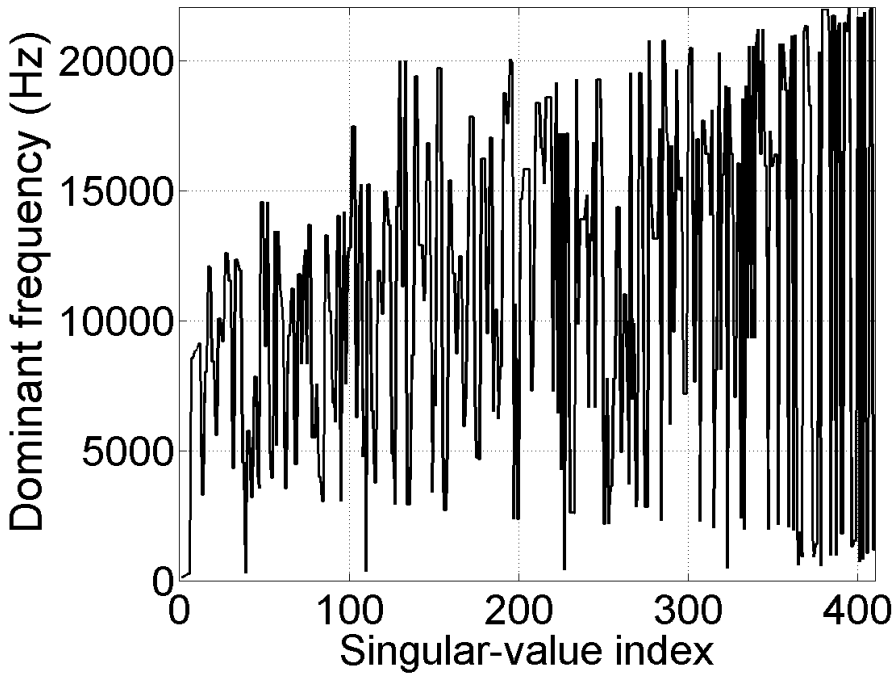

Figure 4.3: Relationship between dominant frequency and singular-value index of the signal (track no. 57).

and from different audio signals are shown in Fig. 4.4. Some share common properties. For example, Figs. 4.4(c), 4.4(e), and 4.4(f) look more like the relationship of which the frequency increases as the index increases. Figures 4.4(a) and 4.4(b) have more oscillation at the higher indices than at the lower indices. The relationship in Fig. 4.4(d) looks almost like random.

This proposed scheme makes the assumption that the frame to be embedded the watermark has the relationship between the singular-value index and the peak frequency as those in Figs. 4.4(c), 4.4(e), and 4.4(f). Actually, those in Figs. 4.4(a) and 4.4(b) are also acceptable because our proposed scheme has never embedded the watermark into the indices greater than $\frac{L}{2}$, where L is the window length for the matrix formation of the basic SSA. To characterize these relationships, we first define the degree of dispersion, ζ, of the relationship curve as follows.

$$\zeta = \sum_{i=1}^{\frac{L}{2}} \left| f_p(i) - \Theta(i) \right|. \tag{4.2}$$

$$\Theta(i) = \left(\frac{f_p(\frac{L}{2}) - 1}{\frac{L}{2} - f_p(1)} \right) \times \left(i - \frac{L}{2} \right) + f_p\left(\frac{L}{2} \right), \tag{4.3}$$

where i is the singular-value index, $f_p(i)$ is the peak frequency of the oscillatory component associated with the index i, $\Theta(i)$ is the line connecting two points $\big(1, f_p(1)\big)$ and $\Big(\frac{L}{2}, f_p(\frac{L}{2})\Big)$, and L is the window length for the matrix formation.

Then, we define ζ_{ref} as the degree of dispersion of the reference curve, as shown in Fig. 4.5, of which the differences between $f_p(i)$ and $\Theta(i)$ for each $i = 1$ to L are set to 2 kHz. Finally, the degree of dispersion ζ_{B} (in Bel) of the relationship curve, compared with the reference curve, is defined as

$$\zeta_{\text{B}} = \log \frac{\zeta}{\zeta_{\text{ref}}}. \tag{4.4}$$

The value of ζ_{B} can be used to characterize the relationships between the peak frequency and the singular-value index, i.e., the lower the value of ζ_{B}, the curve the more like the reference curve. Examples of the distribution of ζ_{B} over 100 frames of two audio signals are shown in Fig. 4.6. Based on our observation, the curve with the value of ζ_{B} lower than 0.5 Bel looks more like the curves in Figs.4.4(a), 4.4(b), 4.4(c), 4.4(e), and 4.4(f), and we found this kind of curves about half the time, for instance, 45% in the signal 1 and 77% in the signal 2 of Fig. 4.6.

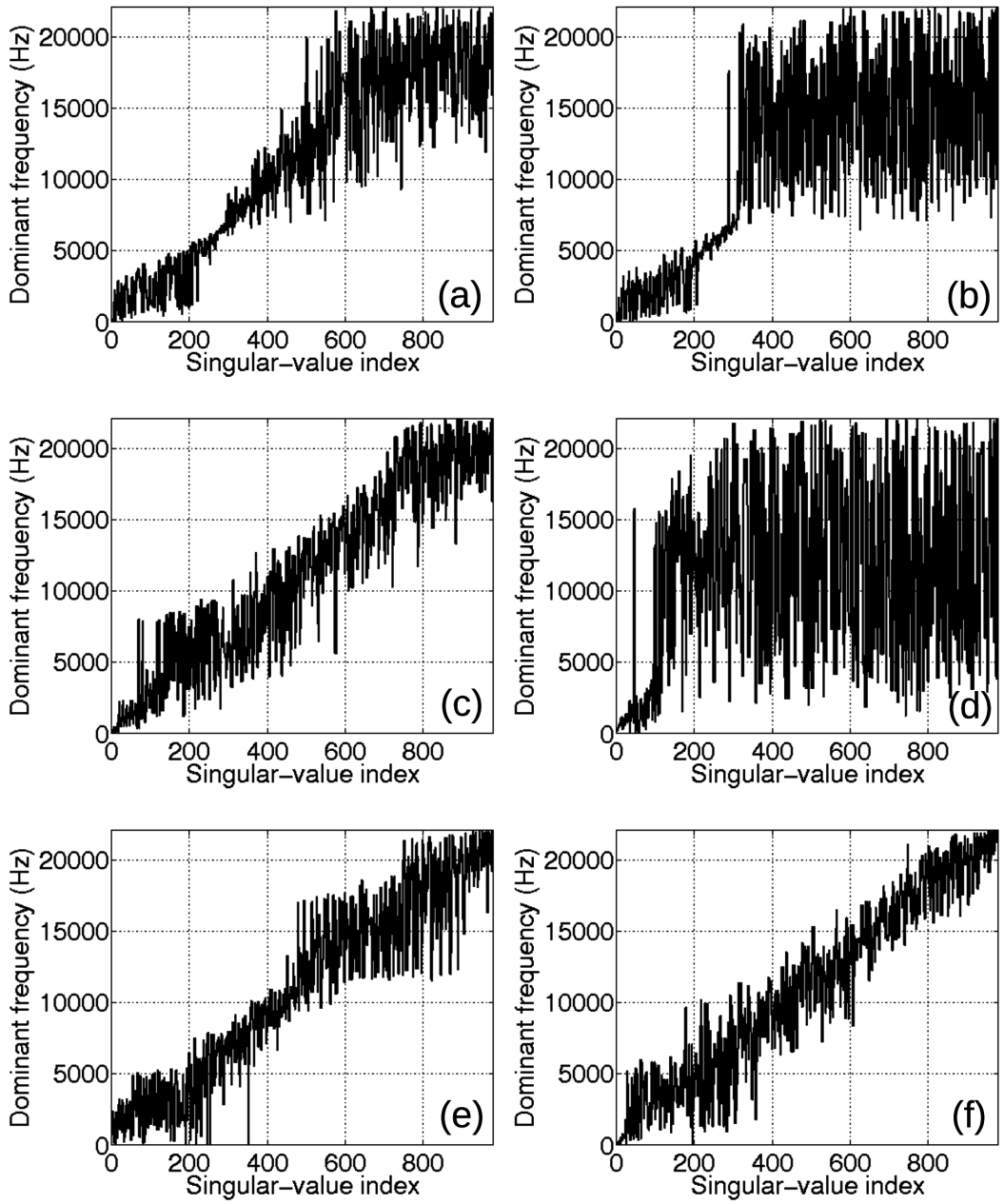

Figure 4.4: Relationships between the dominant frequency and the singular-value index from six different frames.

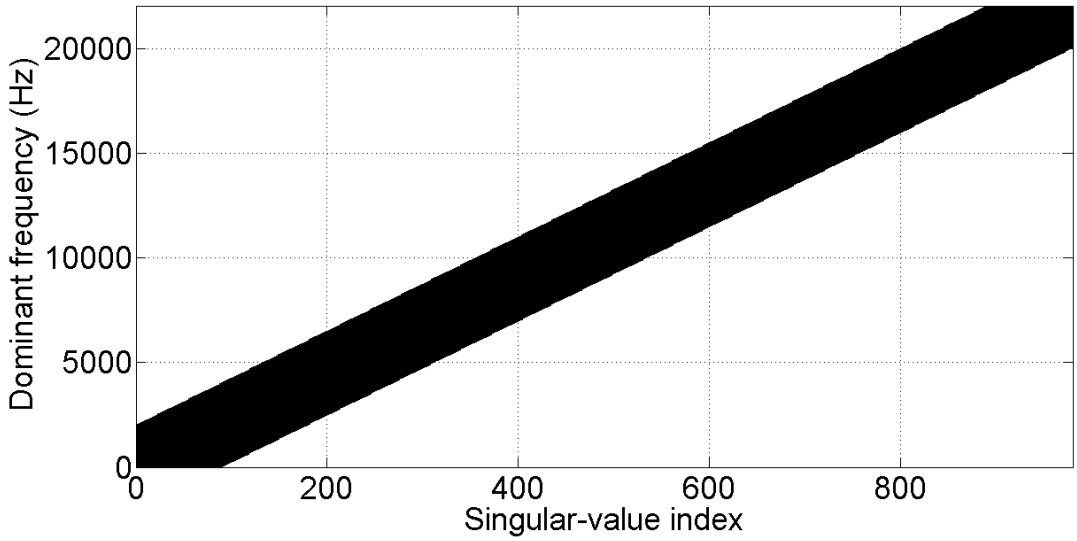

Figure 4.5: Reference relationship between the dominant frequency and the singular-value index.

This finding implies that, at least, those frames with the $\zeta_{\mathrm{B}} < 0.5$ Bel will be improved considerably in terms of the inaudibility when we apply the proposed method, which is based on the psychoacoustic model and this kind of the index-frequency relationship, to determine the parameters u and l.

4.4 Experiments on the automatic frame detection

The SSA-based AIH scheme with the automatic frame detection was implemented based on the scheme described in Sect. 3.5.2 with the parameters shown in Table 4.20.

To evaluate the scheme, 30 audio signals were randomly chosen from the RWC database. Each signal is clipped to 48000 samples randomly. Then, 2 to 7 locations were randomly generated for embedding watermark bits. There were 140 bits in total.

The accuracy of our automatic frame detection was 80%. (112 out of 140 clips were correctly detected.) The false positive detection was 6.42%. The sound quality evaluations are shown in Figs.4.7, 4.8, and 4.9. The average ODG, LSD, and SDR were 0.05, 0.17 dB, and 32.56 dB, respectively.

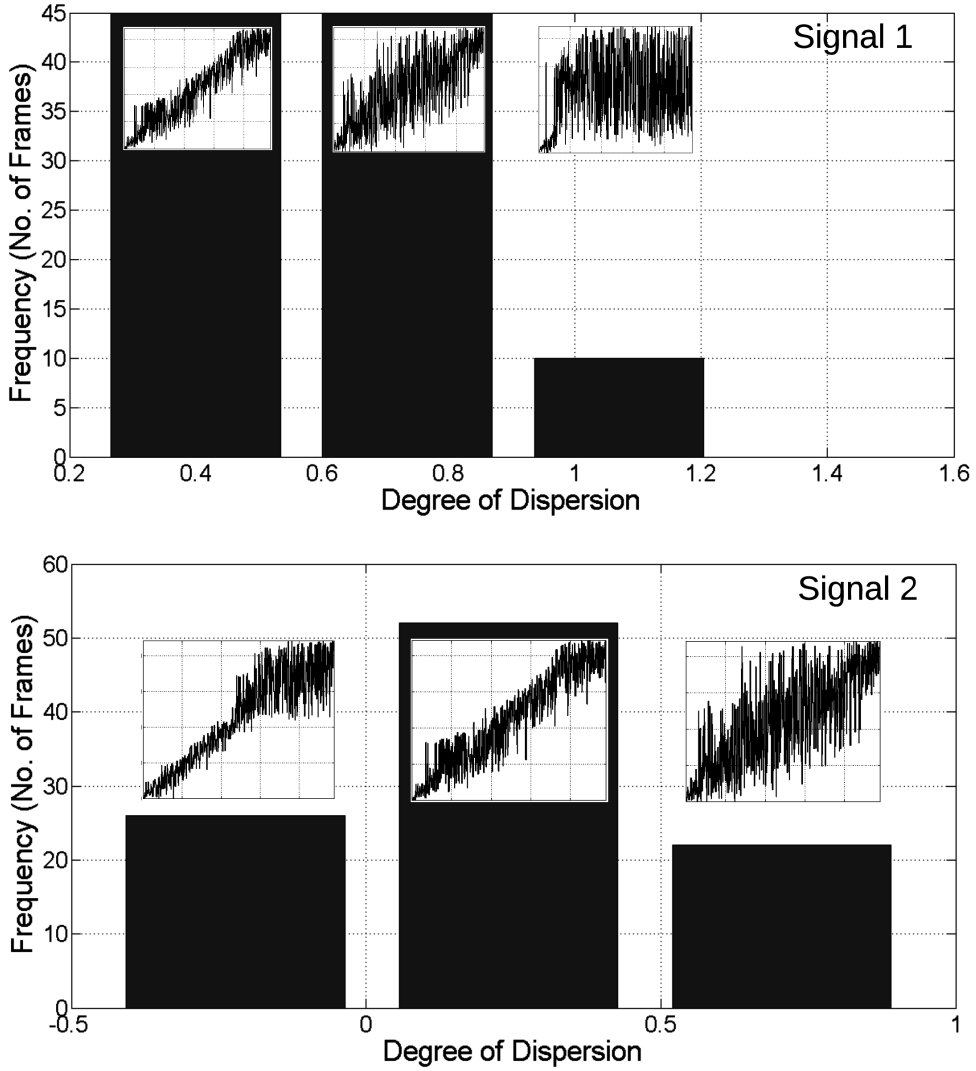

Figure 4.6: Examples of the distribution of ζ_B over 100 frames of two audio signals.

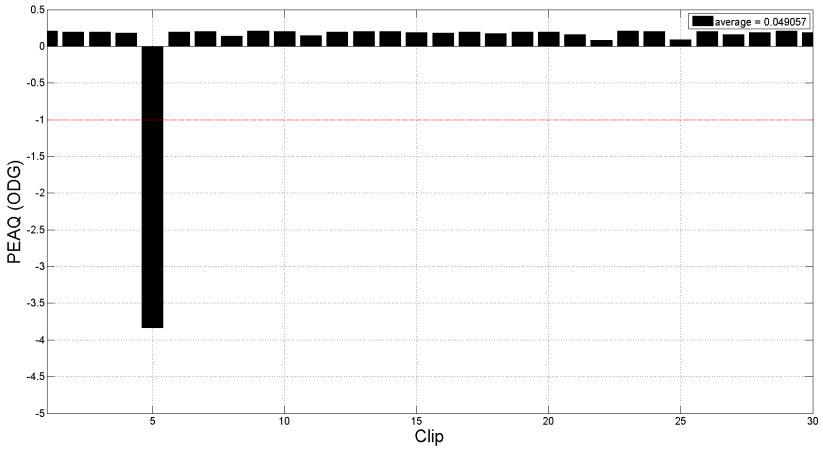

Figure 4.7: ODGs of watermarked signals obtained from the scheme with the automatic frame detection.

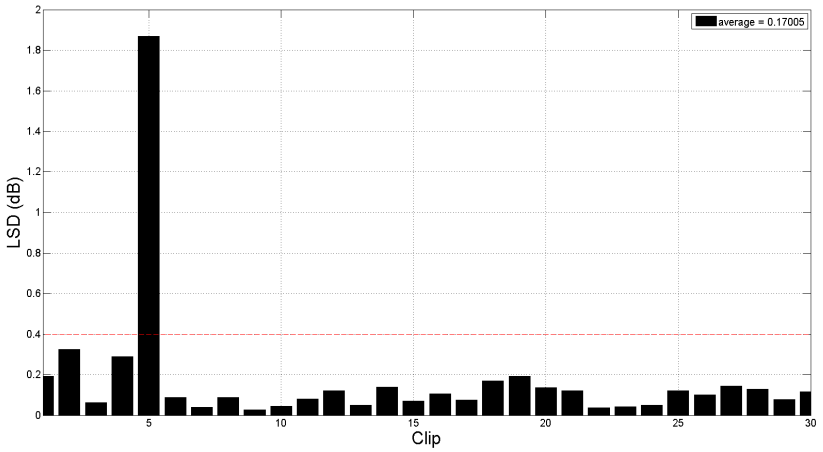

Figure 4.8: LSDs of watermarked signals obtained from the scheme with the automatic frame detection.

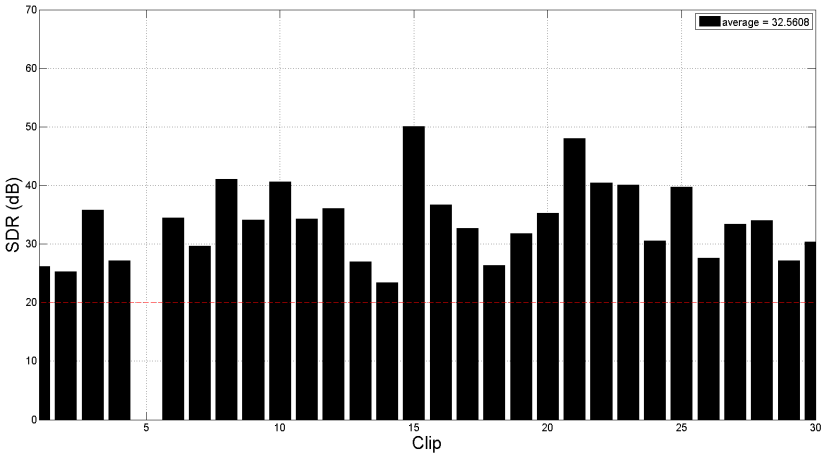

Figure 4.9: SDRs of watermarked signals obtained from the scheme with the automatic frame detection.

Table 4.20: Parameters for the SSA-based AIH scheme with the automatic frame detection.

Subframe length N (sample)	800
Window length L (sample)	320
Subframe-scan step size δ	10
Frame-scan step size Δ	10
Overlap margin Σ	20
Maximum embedding capacity (bps)	13.78
Payload (bit/audio signal)	2-7
Total duration (second)	1.1

Discussion

The sound quality is very good even though the subframe size is very small. There are two facts that can explain these results. First, the capacity is four-time less than that of the scheme without the automatic frame detection since four subframes are required to represent one watermark bit. Second, singular values are disturbed only when subframes are embedded the watermark bit 1 and are untouched when embedding the watermark bit 0.

The detection rate of 80% can confirm the fundamental concepts on which the proposed scheme is based. From our analysis of data, we found that the detection rate is determined by the algorithm that interprets the bit string b_i. In the proposed scheme, our algorithm uses the simplest rectangular windows to find the pattern of b_i. Even in the case the algorithm could not detect a watermark bit, we found that the string b_i correctly presented the concavity on singular spectra, but with distortion to some degree. Therefore, effective pattern recognition techniques could be helpful in this situation.

Unexpectedly, false positive detections occur, i.e., the algorithm reports that there is a watermark when there is no hidden information embedded there. We investigated this problem by analyzing unwatermarked signals with the proposed automatic frame detection. We found that in those false-positive-detection cases, there are natural concavities on singular spectra. If the false positive detection is a serious concern, we can get around the problem by detecting the natural concavity and then hiding information only in the

Table 4.21: Parameters for the AIH scheme based on SSA in DWT domain.

Frame length N (sample)	2448
No. of the filter bank levels	3
Wavelet name	Daubechies 1 [143]
No. of the detail coefficients	2142
Window length L (sample)	856
Parameter u	5
Parameter l	100
Embedding capacity (bps)	18
Payload (bit)	150
Total duration (second)	8.5
Repetition	No

no-concavity frames. Otherwise, a good pattern recognition is required due to our findings that the patterns of the string b_i of the natural concavity are different from those of the embedded watermark. This problem will be investigated in the future.

4.5 Experiments on AIH based on SSA in DWT domain

The parameters for implementation are shown in Table 4.21. We chose the 3-level DWT with the wavelet Daubechies 1. The frame size N is a bit smaller than that of the fixed-parameter model because the n-level DWT requires that N is divisible by 2^n, and we have $N - \frac{N}{2^n}$ detail coefficients in total. The trajectory matrix is constructed by using a sequence of detail coefficients ranging from low to high frequencies, i.e., from the level 3 to the level 1.

Evaluation and discussion

Compared with the SSA-based AIH in time domain, the performance in terms of robustness and inaudibility of the scheme in DWT domain is rather poor. The ODG, LSD, and SDR are -0.4429, 1.5063 dB, and 23.4094 dB, respectively. The average BER is 5.74%, and it increases to around the chance level when attacks were applied except for the Gaussian noise addition.

There are many opened questions at this point. For example, what will happen when we form the trajectory matrix differently, when we change the mother wavelet, or when we increase or decrease the filter bank levels. We leave these curiosity for future work. Besides, the myth about the singular value loses its physical meaning when a trajectory matrix is formed by wavelet coefficients. Although the idea of embedding information in high-frequency coefficients seems reasonable, it is not clear what are singular values of the matrix formed by them.

4.6 Summary

Compared with the conventional method, the fixed-parameter model was more robust. However, the sound quality of watermarked signals obtained from this model was poorer. We found that the frame size affected both the sound quality of watermarked signal and the robustness of the watermark. The bigger size is more robust and inaudible than the smaller one. Embedding repetitions can reduce the BER, but, at the same embedding capacity, the bigger frame size is still better in robustness. There was no significant difference between extractions by the median and by the polynomial fitting for the fixed-parameter model.

In comparison with the fixed-parameter model, the partially-blind model outperformed it in terms of the sound quality. The sound-quality measures of watermark signals obtained from the partially-blind model and the conventional method were comparable, but the robustness of the partially-blind model was significantly better. The robustness of the partially-blind model dropped considerably when compared with that of the fixed-parameter one. Therefore, there was clearly a trade-off between inaudibility and robustness. The completely-blind model could extract a watermark blindly, but its detection rate decreased due to the blindness condition.

When the psychoacoustic model was integrated into the scheme, the sound quality of watermarked signal was the best. The automatic frame detection could detect frames with the accuracy of 80% and the sound quality of watermarked signals was very good. The scheme based on SSA in DWT domain was the most fragile, and the singular values of this scheme lost their physical meaning.

Chapter 5

Applications of the SSA-based AIH

Three applications of the SSA-based AIH are proposed in this chapter: ownership protection, information carrier, and fragile audio-watermarking. Each application has its own requirements, and some of their requirements are different. For example, the fragile audio-watermarking requires fragility, but the ownership protection requires robustness. However, the robustness is not a concern for the information carrier. These applications are a good example that shows the flexibility of the proposed SSA-based AIH. Evaluations with respect to each application's requirements are performed, and results are reported.

5.1 Ownership protection with SSA-based audio watermarking

When there is disagreement over ownership of audio work, the watermark can serve the useful purpose of identifying the true owner. Only the owner who knows the secret message can reveal the existence of it. The important properties that the AIH scheme for the ownership protection or ownership identification requires are the secrecy of the watermark and robustness. We have already shown that all proposed models are robust against many signal processing attacks. Due to the general requirement of inaudibility, the audio watermark is a secret message to some degree if there is no explicit expression somewhere stating that its host is watermarked. Straightforwardly, to increase the degree of secrecy, any efficient encryption technique can be integrated to the watermarking scheme [144–146].

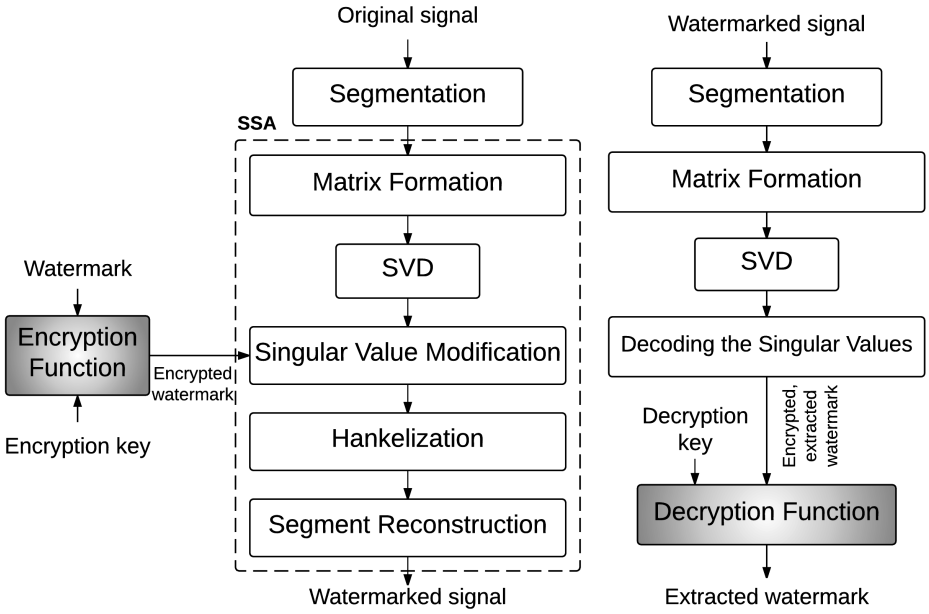

Figure 5.1: Embedding and extraction processes of the AIH scheme for the ownership protection.

5.1.1 Implementation

All models proposed in Chapter 3 can be directly applied to this application. In this demonstration, we show an example of applying the most widely used RSA (Rivest-Shamir-Adleman) algorithm [147] to the fixed-parameter model. The simplest model is chosen in order to reduce the computational time in our simulation. It is obvious that the better performance in inaudibility could be achieved if the adaptive-parameter models are the main structure of this application.

The embedding and extraction processes are shown in Fig. 5.1 and similar to those of the fixed-parameter model in Sect. 4.2.1. There are two differences between Fig. 5.1 and Fig. 3.1. First, instead of embedding the watermark, the embedding process in Fig. 5.1 embeds the encrypted watermark. Second, after decoding singular values, the extraction process in Fig. 5.1 has to decrypt the encrypted data to deliver the extracted watermark. Thus, the additional processes are the encryption with the encryption key (public key) and decryption with the decryption key (private key).

The concept of the RSA encryption and decryption is quite simple [148–150]: an integer m, such that $0 \leq m < n$ and $\gcd(m, n) = 1$, where gcd is the greatest common

divisor, is encrypted to the integer $c \equiv m^e \pmod{n}$, where (e, n) is the encryption key. The decryption key (d, n) is used to decrypt the message c by computing $c^d \equiv (m^e)^d \equiv m (\mathrm{mod} n)$. The RSA cryptosystem generates a pair of public- and private-keys by the following steps.

1. Compute an integer n as the product of two very large, random primes p and q.
2. Select an integer e, which is an odd number between 3 and $n-1$, that is a relatively prime to $p-1$ and $q-1$.
3. Pick an integer d to be a large number by random, which satisfies the condition $\gcd(d, (p-1) \times (q-1)) = 1$.

5.1.2 Evaluation and discussion

In information hiding sense, there is no difference between this application and the implementation of the fixed-parameter model since the only difference between them is the contents of the watermark. Therefore, the sound quality is at the same level. Actually, the robustness is at the same level as well if we calculate the BER by comparing the encrypted watermark with the extracted, encrypted watermark. However, from our experiment with 3 different watermark bit-strings embedded into one audio signal (track number 1) from the RWC database, the average BER increases to about 50% when we compare the watermark with the extracted watermark in the case of wrong decryption keys. This result is to be expected anyway.

In this application, we show how to apply the proposed scheme without any modification. In the next two applications, we will show how to modify the proposed scheme to meet the requirements of other applications.

5.2 Information carrier with SSA-based audio watermarking

In this application, the watermark is the secondary information that adds values to the host signal. There is no motivation for attackers to attack or destroy the watermark [14]. Thus, the robustness is not required. Two mandatory requirements are blind detection and high capacity.

The embedding capacity of the SSA-based AIH scheme is associated with the frame size. The higher capacity means the smaller frame size, and, as a result, the fewer singular values. When the singular values are too few, it is difficult to select the embedding area (specifically, the parameters u and l) without adversely affecting the sound quality. This is because, according to the proposed singular-value modification rule, all singular values of the interval $[u+1, l-1]$ are rounded up or down to the upper- or lower-bound values, respectively. However, the SSA-based AIH scheme can still achieve the high capacity by adopting the quantization index modulation (QIM) [151–153], i.e., all singular values are quantized by the QIM technique.

5.2.1 Implementation

The implementation of this application is similar to the implementation of the fixed-parameter model. The embedding and extraction processes can be depicted as Fig. 3.1 on the left and the right, respectively. The differences are in the steps of the singular-value modification and decoding the singular value. We applied QIM steps as described by Vivekananda Bhat K. et al. [23] to the main scheme.

Embedding process

The embedding process consists of six steps as follows.

1. The host signal is segmented into non-overlapping frames.
2. Each frame is represented by the trajectory matrix.
3. Each matrix is factorized by SVD to deliver the singular values.
4. All singular values obtained from the previous step are quantized by the following procedure.
    - Given a singular spectrum $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_n}\}$, the Euclidean norm of the singular spectrum is defined as
    $$\left\|\sqrt{\lambda}\right\| = \sqrt{\sqrt{\lambda_1}^2 + \sqrt{\lambda_2}^2 + \cdots + \sqrt{\lambda_n}^2}. \tag{5.1}$$
    - All singular values $\sqrt{\lambda_i}$ for $i = 1$ to n are multiplied by the factor of $\frac{\left\|\sqrt{\lambda}_w\right\|}{\left\|\sqrt{\lambda}\right\|}$,

where $\left\|\sqrt{\lambda}_w\right\|$ is a function of the watermark bit $w \in \{0, 1\}$.

$$\left\|\sqrt{\lambda_w}\right\| = \begin{cases} \Delta \cdot \left(\lfloor \frac{\left\|\sqrt{\lambda}\right\|}{\Delta} \rfloor + d - (\lfloor \frac{\left\|\sqrt{\lambda}\right\|}{\Delta} \rfloor \bmod 2) \right) + \frac{\Delta}{2}, & \text{if } w = 0 \\ \Delta \cdot \left(\lfloor \frac{\left\|\sqrt{\lambda}\right\|}{\Delta} \rfloor + d - \left((\lfloor \frac{\left\|\sqrt{\lambda}\right\|}{\Delta} \rfloor + d) \bmod 2 \right) \right) + \frac{\Delta}{2}, & \text{if } w = 1, \end{cases} \tag{5.2}$$

where d and Δ are a user-defined parameter and the quantization step, respectively. In our experiment, we set $d=1$ and $\Delta=0.5$.

5. After replacing old singular values with the new ones, each matrix is mapped to the watermarked frame by Hankelizing.
6. Finally, the watermarked signal is constructed by combining those frames.

Extraction process

The first three steps of the extraction process are exactly the same as those of the embedding one. After obtaining the singular spectrum $\{\sqrt{\lambda_1}, \sqrt{\lambda_2}, ..., \sqrt{\lambda_d}\}$, the extracted watermark-bit $\hat{w}$ is 1 if $(\lfloor \frac{\left\|\sqrt{\lambda}\right\|}{\Delta} \rfloor \bmod 2)$ is equal to 0; otherwise, the extracted watermark-bit $\hat{w}$ is 0.

5.2.2 Evaluation

Our implementation was evaluated at 16, 32, 64, 128, 256, and 512 bps. Each bit-rate relates to the frame sizes of 2756, 1378, 689, 344, 172, and 86 samples, respectively.

The relation between the BER and the embedding capacity is shown in Fig. 5.2. At the capacity of 512 bps, the BER is still lower than 10%. However, the sound quality decreases as the capacity increases, as shown in Figs. 5.3, 5.4, and 5.5.

5.2.3 Discussion

Although all the three sound-quality measures (Figs.5.3, 5.4, and 5.5) agree that sound-quality distortion in the watermarked signals increases with the capacity, it does not clear which measure is more reliable than the other. For example, according to the IHC's criteria [154], the ODG greater than -2 is acceptable. With 10% BER allowed,

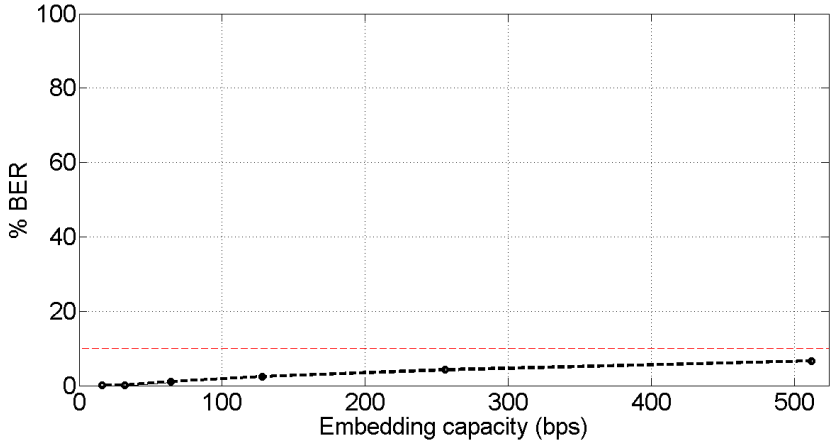

Figure 5.2: Relation between robustness and embedding capacity of the proposed framework for the information carrier.

the embedding capacity of the scheme is at least 512 bps. Differently, based on our observations, if the LSD is greater than 0.4 dB, the distortion can be perceived easily. The capacity is then limited around 100 bps. In speech signal processing, the signal-to-noise ratio (SNR) of which greater than 25 dB indicates a very good signal, and the SDR is similar to SNR. Thus, the maximum capacity is about 170 bps. Even with the smallest embedding capacity, for example, 100 bps, it is still high compared with the fixed-parameter model. At the same level of LSD, the capacity of this scheme is greater than that of the fixed-parameter model at least two times.

Note that the parameters d and Δ was set to 1 and 0.5, respectively, as suggested by Vivekananda Bhat K. [23]. Undoubtedly, those parameters affect sound quality and robustness because they determine the values of modified singular values. Therefore, overall improvement can be achieved by searching for the optimum d and Δ.

5.3 Fragile audio-watermarking based on SSA

Fragile audio watermarking can be used to ensure the credibility of the host signals [155] and to prove the integrity of the content [156]. The watermark is used to verify that the host signal has not been edited since it was embedded. Therefore, it must be fragile to the signal processing. As discussed in Sect. 3.2.2, the SSA-based AIH scheme provides a good insight about the location of embedding areas where we can expect the watermark to be

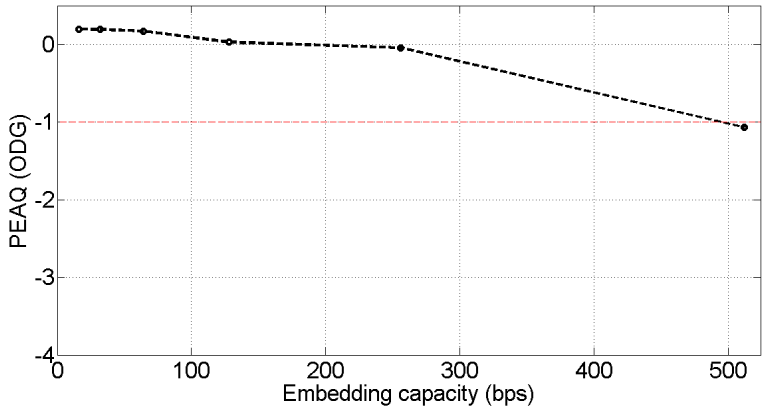

Figure 5.3: Relation between PEAQ and embedding capacity.

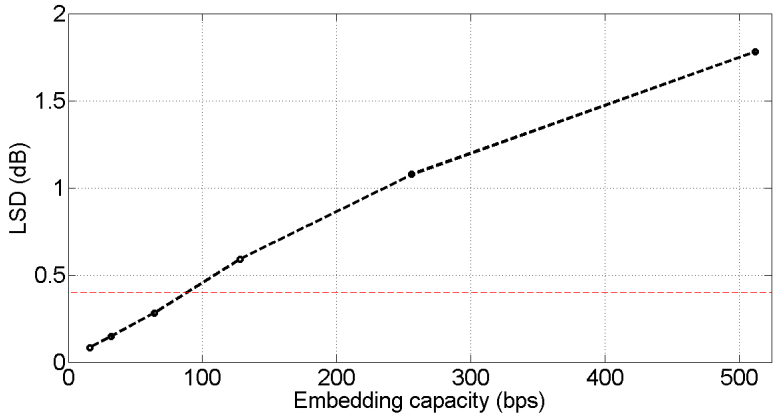

Figure 5.4: Relation between LSD and embedding capacity.

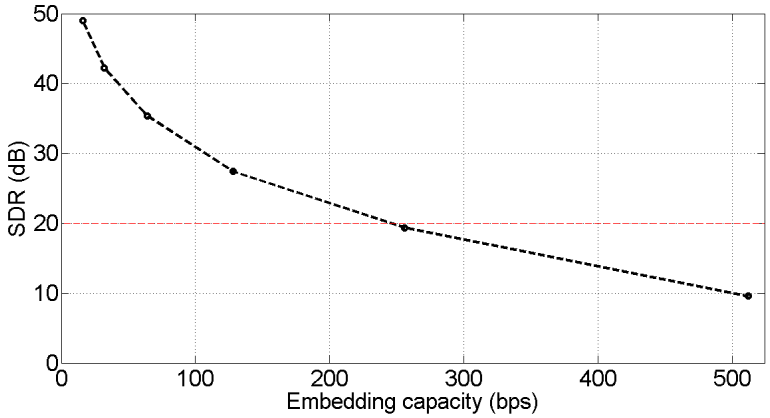

Figure 5.5: Relation between SDR and embedding capacity.

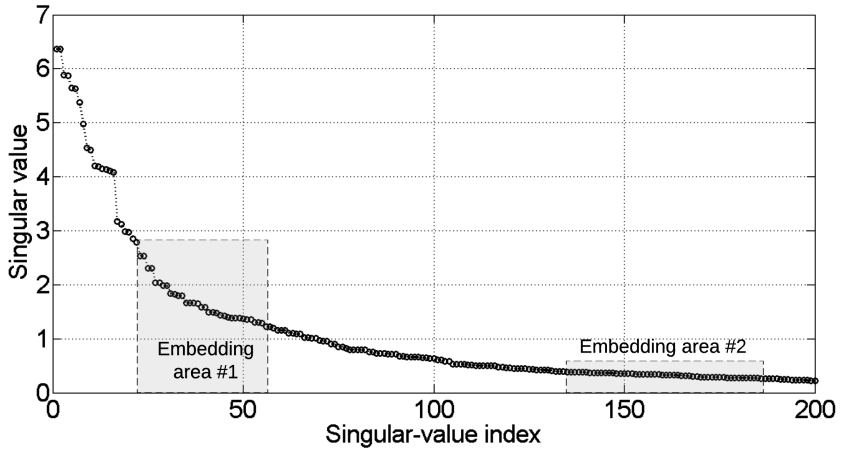

Figure 5.6: Double embedding: information is embedded twice into two areas.

fragile. As shown in Fig. 3.8, the strain increases as the singular-value index increases when common signal processing is performed to the watermarked sound. Therefore, high-order singular values are good candidates for the fragile watermarking.

5.3.1 Implementation

The implementation of this application can be done easily by applying the core structure of the SSA-based AIH to embed the watermark twice. The first embedding performs on the low-order singular values as the fixed-parameter does, whereas the second embedding performs on higher-order ones, as demonstrated in Fig. 5.6. Then, the integrity of watermarked signals can be verified by the watermarks extracted from both areas.

Embedding process

The embedding process consists of seven steps as follows.

1. The host signal is segmented into non-overlapping frames.
2. Each frame is represented by the trajectory matrix.
3. Each matrix is factorized by SVD to deliver the singular values.
4. Two intervals of singular values $[u_1, l_1]$ and $[u_2, l_2]$ are chosen, and the equation 3.1 is adopted to modify the singular values within these two intervals.
5. After the singular-value modification, each matrix is mapped to the watermarked frame by Hankelizing.

6. Finally, the watermarked signal is constructed by combining those frames.

Extraction process

The extraction process described in Sect. 3.2.3 is performed to the watermarked signal twice in order to extracted two watermarks $\hat{w_1}(i)$ and $\hat{w_2}(i)$ for $i=1$ to N. We can use only the BER of the second extracted the watermark $\hat{w_2}(i)$ or the difference between the BERs of these two watermarks to verify the genuineness of the signal carried the watermarked. For example, if the difference is greater than a predefined value, the watermarked signal is tampered.

5.3.2 Evaluation

In our experiment, two intervals were from indices 20 to 60 and indices 100 to 200. The frame size was 1632 samples, and 150 bits of watermark were embedded into 100 audio signals. The aim of this experiment is just to show an effect on the watermarks when watermarked signals are modified, and we can use that effect to detect tampering.

The BER comparisons are shown in Figs. 5.7 and 5.8 in the cases of no attack and MP3 attack, respectively. The average LSD is 0.6085 dB.

5.3.3 Discussion

It can be seen that MP3 compression damages the second watermark more severely. However, in this implementation, the detection rate when there is no attack is also not good. If we do not know the expected difference between BERs in the case of no attack in advance, it may be difficult to judge whether an audio signal is tampered. One possible solution is by selecting the proper index-intervals for embedding the watermarks. The differential evolution can be applied similarly to the adaptive-parameter models.

5.4 Summary

The proposed schemes were applied in three applications. For the ownership protection, the fixed-parameter model was integrated with the RSA algorithm to protect hidden information. Without the decryption key, the average BER increased to about 50%. For

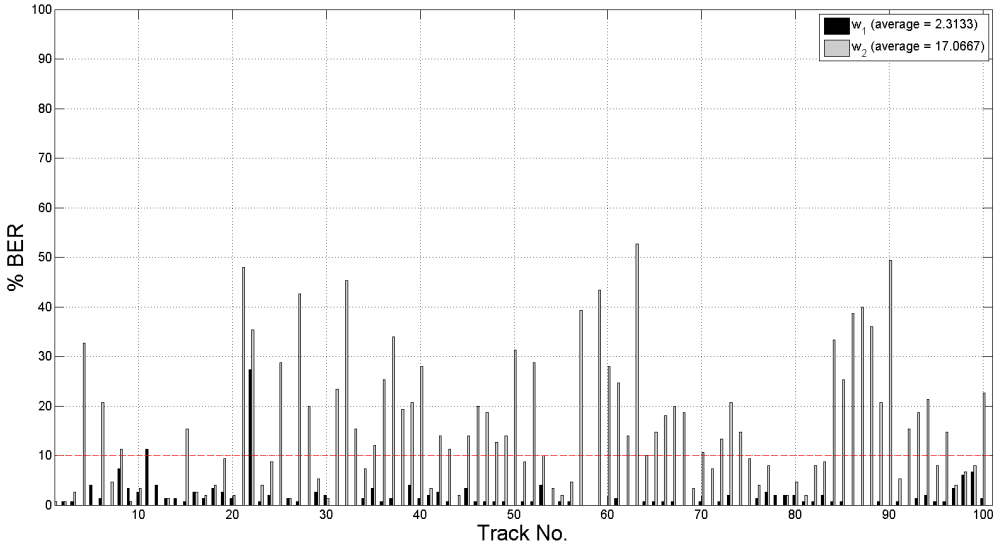

Figure 5.7: BERs (%) of double embedding when no attack.

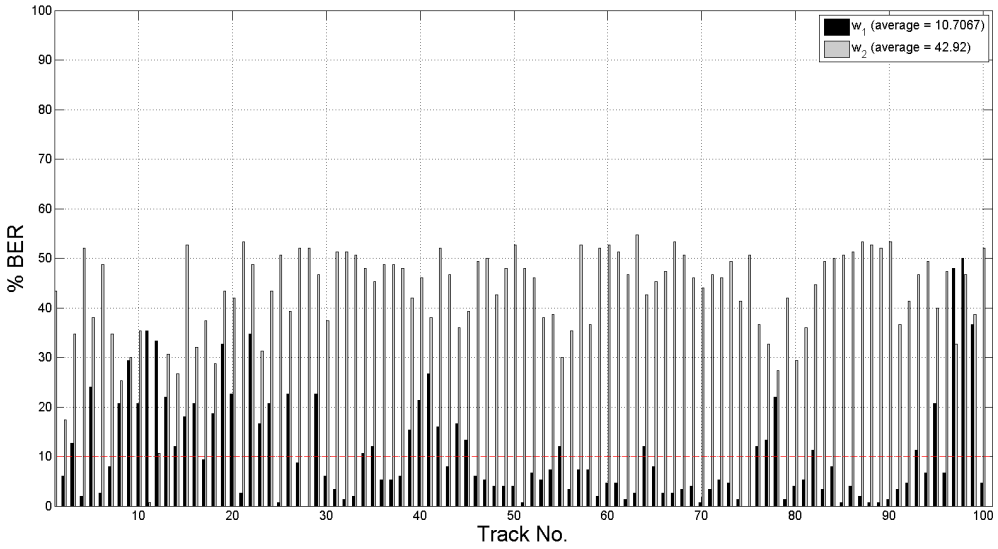

Figure 5.8: BERs (%) of double embedding when MP3 was performed to watermarked signals.

the information carrier application, the core structure was modified to embed a watermark bit into all singular values by using QIM. The embedding capacity was increased up to 170 bps at the SDR of 25 dB. For the fragile audio-watermarking, a watermark bit was embedded into the high-order singular values which have a high strain. Thus, the watermark bit was destroyed easily.

The performances of all three applications could be improved enormously since all parameters have not yet been adjusted to their best values.

Chapter 6

Conclusion

This chapter summarizes this work and emphasizes its contributions to AIH research field as well as to other research fields. Since the ultimate goal of audio information hiding has yet to achieve, it discusses room for improvement in the last section.

6.1 Summary

In this work, we used SSA to analyze signals and investigated the characteristic of singular spectrum in order to propose a novel AIH framework. To be concise, let us break down the basic findings from our study first.

1. A singular value is slightly changed under many signal processing attacks. This property makes it be a good choice for hiding information.
2. A singular value delivered by SSA can be interpreted as a scale factor of an oscillatory component. Since singular values are sorted in descending order, the lower-order singular values have contributed more to the signal.
3. A singular spectrum is naturally convex, and we can embed information into it by forcing a small part of it to be concave.
4. A strain is defined to measure the deformation of a singular value when it is under attack. We found that a singular value with a higher order has a higher strain. In other words, the higher-order singular value changes its value easier than the lower-order one. Based on this finding, we can make a general rule for embedding: modifying lower-order singular values makes a watermark more robust, and

modifying higher-order singular values makes a watermark more fragile.

5. We found that modifying higher-order singular values makes a watermark more inaudible. (This fact can be explained by 2.) Therefore, there is a trade-off if we consider 4 and 5 together.
6. Even though the relation between singular-value indices and frequency components of signals is not clear to formulate analytically, there exists a trend, and the relation can be established empirically.

These facts are fundamental to our work and are used to build six SSA-based AIH schemes: the fixed-parameter model, the partially-blind model, the completely-blind model, the AIH based on SSA and the psychoacoustic model, the AIH based on SSA in DWT domain, and the SSA-based AIH scheme with the automatic frame detection. All of them can be considered as variants of the same framework.

Compared with the conventional method and all other models, the fixed-parameter model is most robust. However, the sound quality of watermarked signals obtained from it is poorest. The reason is that this model modifies the lowest-order singular values compared to the other models. The results from the fixed-parameter model can achieve the first subgoal (as stated in Chapter 1) because it verifies that the SSA-based scheme can keep the advantages of the SVD-based technique.

Compared with the fixed-parameter model, the adaptive-parameter models (both completely-blind and partially-blind) outperformed it in terms of the sound quality because the differential evolution optimizer is used to find the balance between inaudibility and robustness. Thus, the success of the adaptive-parameter models can be considered as the achievement in the second subgoal since it shows that the proposed scheme has parameters that can be adjusted to gain better performance. In addition, the robustness of the partially-blind model is better than that of the conventional method.

When the psychoacoustic model is integrated into the scheme, the sound quality of watermarked signal is the best. The results verify the connection between singular spectrum and frequency components of the signal and show that the SMR from the psychoacoustic model can be used to reduce distortion in the sound quality. One drawback of this implementation is that it is fragile to signal processing because the criterion that we used to determine the parameters is too simple.

The proposed automatic frame detection can detect frames with the accuracy of 80% and the sound quality of watermarked signals is very good compared with the fixed-parameter model. The scheme based on SSA in DWT domain is the most fragile, and the singular values of this scheme lost their physical meaning.

Finally, three applications were proposed: ownership protection, information carrier, and fragile audio-watermarking. For the ownership protection, the fixed-parameter model is integrated with the RSA algorithm to protect the hidden information. Without the decryption key, the average BER increased to about 50%. For the information carrier application, the core structure is modified to embed a watermark bit into all singular values by using QIM. The embedding capacity is increased up to 170 bps at the SDR of 25 dB. For the fragile audio-watermarking, a watermark bit is embedded into high-order singular values, which have a high strain. Thus, the watermark bit is destroyed easily.

The performances of all three applications could be improved enormously since all parameters in the implementations were not adjusted to their best values. However, it does not go so far to say that at least five subgoals are checked.

To sum things up, the unique and novel points of this work are as follows.

1. We proposed AIH framework based on SSA because we want to exploit the strength of the SVD-based one. We can deal with drawbacks of the SVD-based framework by using the fact that, based on SSA, we can know the meaning of singular values. In addition, the SSA-based AIH framework by itself has never been proposed before our work.
2. Prior to the present time, all SVD-based schemes have embedded hidden information into host signal by using various uninformed rules. In contrast, this work shows that the human perception model can be incorporated into the basic SSA-based framework so that the embedding rule can be informed. In order to achieve that state, we have bridged the gap between the standard analysis tool and the SSA. The results show that the proposed schemes can improve the sound quality a great deal.
3. We also show that a singular spectrum has a very useful characteristic for information hiding, i.e., its convexity by nature. We can exploit this characteristic for self-synchronization by looking for concavity through the proposed scan operation.

6.2 Contributions

The contributions of this work can be seen from different viewpoints. Broadly speaking, to society, it offers a solution to social concerns (Sect. 1.1) and to applications that require AIH (Sect. 2.1.2).

This research has explored a non-standard signal processing technique based on SVD. At least, it is not a conventional analysis method for audio/speech signal processing. We also investigate the properties and characteristics of the singular values delivered by this method. We discovered that the singular spectrum curve is approximately convex and made good use of it to hide the information. In order to improve the performance, we combined the psychoacoustic model to the SSA-based scheme and made the connection between them. This framework is general and not limited to the AIH applications.

From the engineering viewpoint, this research contributes to the AIH research field by developing a new and effective AIH framework with a few variants. We show their potentiality to solve different problems with different requirements. Moreover, the algorithms that we develop, such as the concavity detection or self-synchronization, are simple and can be used to solve other similar problems in other fields.

Information hiding in audio is the youngest in its family, and some published methods in the audio watermarking originate from other domains. For example, the SVD-based audio watermarking was influenced by image watermarking. Thus, to the related research fields, this work can contribute to other domains as well. It is not so difficult to imagine using image pixels to form a trajectory matrix, and then its singular values are modified in order to hide information. Even though in some other domains, such as the visual domain, the physical meaning of singular values is lost or changed, but our proposed framework is general enough to be applied.

6.3 Future work

Despite the success of our current SSA-based framework, the ultimate goal has yet to reach. There are rooms for further enhancements.

1. The AIH scheme based on SSA and the psychoacoustic model has shown a good sign of progress towards the ultimate goal. For the next step, we need to increase

the robustness. There are many directions we can pursue. For example, if we persist to use the same psychoacoustic model, we have to use the SMR wisely and more efficiently. In the current version, information is embedded into the singular values that associate with the sub-bands that have SMR less than 10 dB. Experimenting more on the threshold of the maximum limit in dB of SMR, the pattern of SMR curve, and adding the two-tone suppression to the psychoacoustic model and embed information into the suppressed areas that are above the global masking threshold are our future plan. If such the suppressed exist, embedding information there might achieve both inaudibility and robustness. We believe it will achieve inaudibility because of the suppression, and robutness because the frequency range is above the masking threshold.

2. The concept behind the AIH scheme based on SSA in DWT domain seems to be justifiable. If we are insensitive to high-frequency signals, modifying detail coefficients should not strongly affect the sound quality of watermarked signals. Thus, we can modify low-order singular values to gain the robustness. Although the preliminary results show negative, there are a lot of opened questions (as discussed in Sect. 4.5). It might be worthwhile to explore.
3. There are two critical problems for the automatic frame detection. First, the algorithm is time-consuming. Reducing the computational time is one of the future work. At least, two parameters δ and Δ connect with a total number of SVD operations. Increasing them reduces the number of SVD operations, but it surely affects detection accuracy. Then, what are the maximum values of them? Are there any other ways to reduce the time? This is one of the future work. Second, the algorithm used to interpret the pattern of b_i can be improved. Then, it is expected that the detection rate increases.
4. As shown in Eq. 2.1, a trajectory matrix is formed by lagged vectors, and $L-2$ members of two consecutive lagged vectors are overlapping. The interesting question is what happen to the relation between singular-value indices and frequency components if the number of overlapped members is less than $L-2$. Can we exploit it?
5. All three proposed applications can be improved by fine-tuning their parameters.

Appendices

Appendix A

Evaluation details of the proposed SSA-based AIH

One hundred host signals from the RWC music-genre database [133] were used in our experiments. All have a sampling rate of 44.1 kHz, 16-bit quantization, and two channels. A watermark was embedded in one channel. The watermark was embedded starting from the initial segment of host signals. All simulations were operated in the Fujitsu CX250 Cluster (JAIST parallel computers), where each computing node is equipped with two Intel Xeon E5-2680v2 2.80 GHz (10 cores each) and 64 GB memory (4 GB DDR3-1866 ECC $\times$16).

Three distance measures, which are the PEAQ, the LSD, and the SDR, were used to evaluate the sound quality of the watermarked signals.

Five attacks were performed on watermarked signals: Gaussian-noise addition with average signal-to-nois ratio (SNR) of 36 dB, re-sampling with 16 and 22.05 kHz, band-pass filtering with 100-6000 Hz and -12 dB/Oct, MP3 compression with 128 kbps joint stereo, and MP4 compression with 96 kbps.

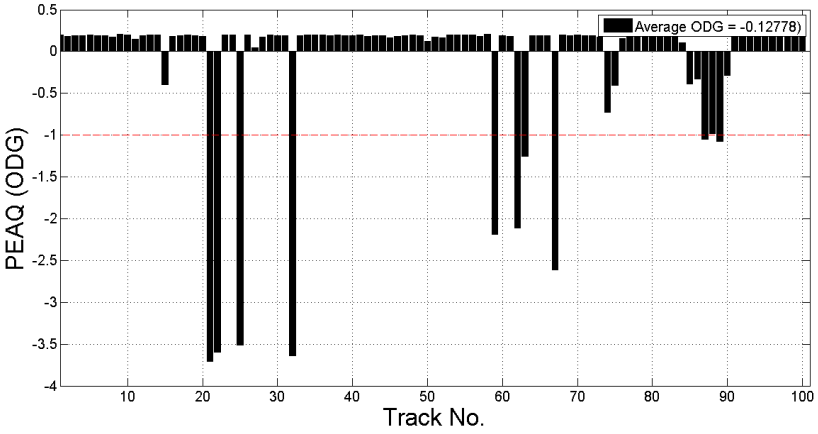

Figure A.1: PEAQ ($N = 2450$, Fixed-parameter model, No repetition.)

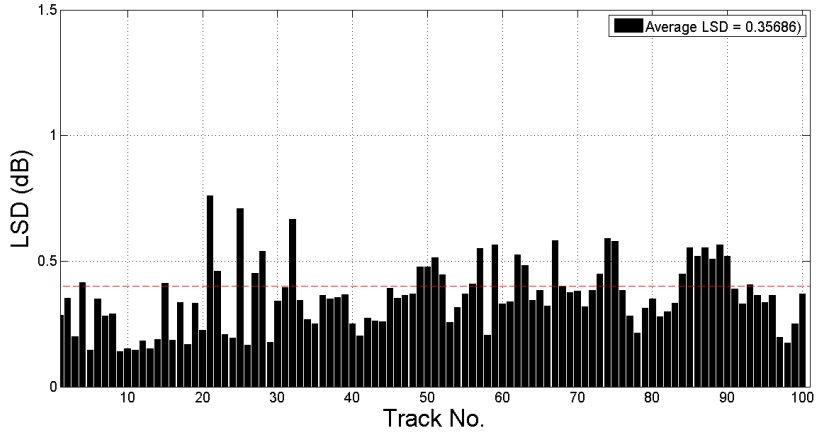

Figure A.2: LSD ($N = 2450$, Fixed-parameter model, No repetition.)

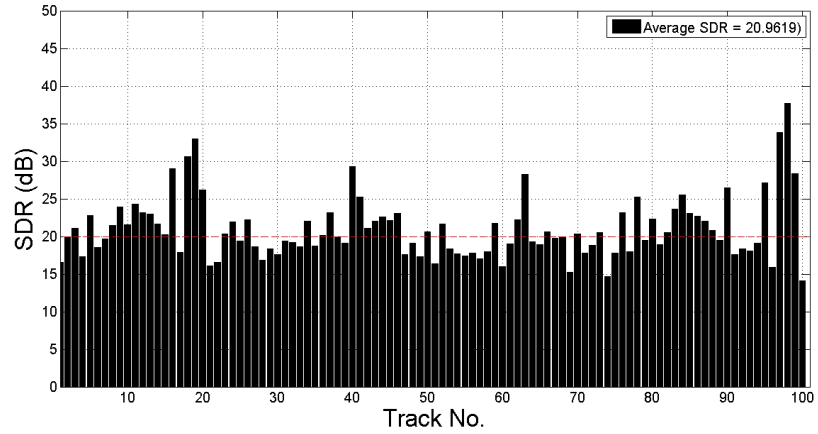

Figure A.3: SDR ($N = 2450$, Fixed-parameter model, No repetition.)

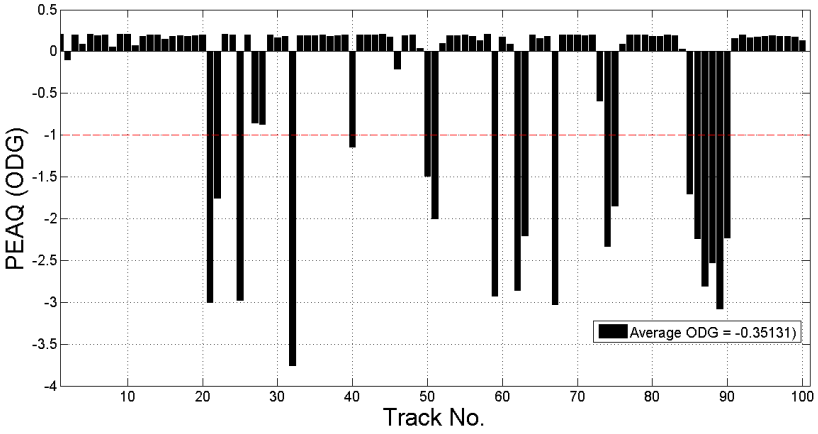

Figure A.4: PEAQ ($N = 816$, Fixed-parameter model, No repetition.)

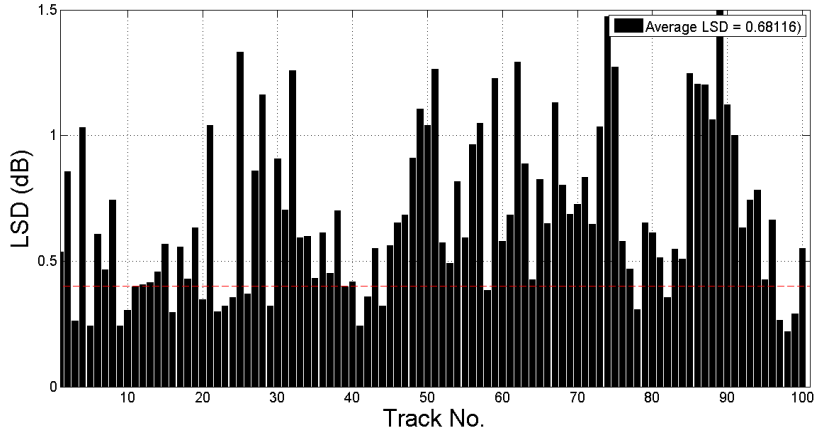

Figure A.5: LSD ($N = 816$, Fixed-parameter model, No repetition.)

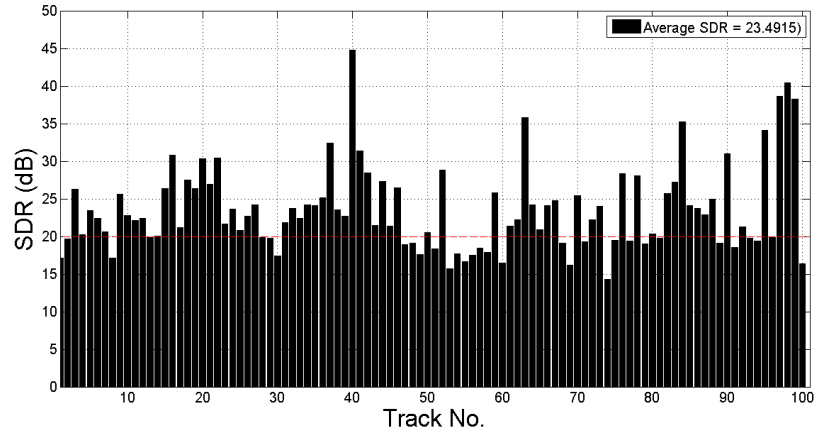

Figure A.6: SDR ($N = 816$, Fixed-parameter model, No repetition.)

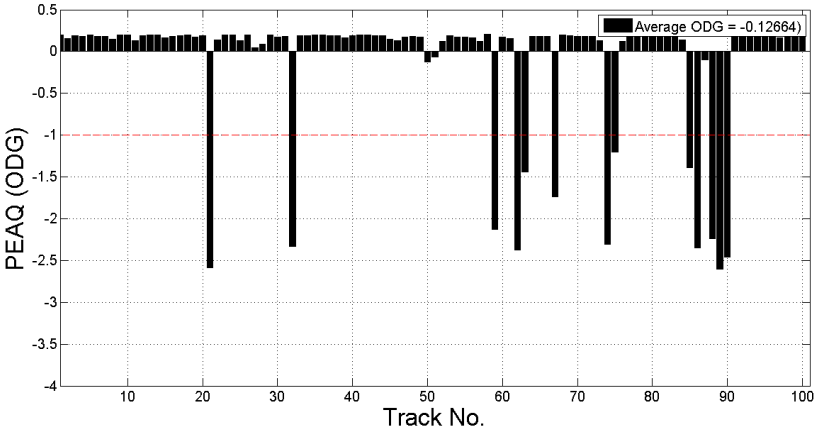

Figure A.7: PEAQ ($N = 816$, Fixed-parameter model, Repetitions $= 5$.)

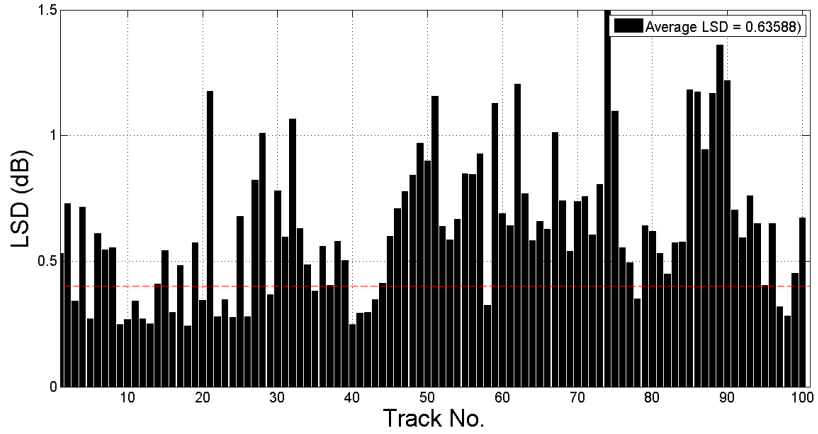

Figure A.8: LSD ($N = 816$, Fixed-parameter model, Repetitions $= 5$.)

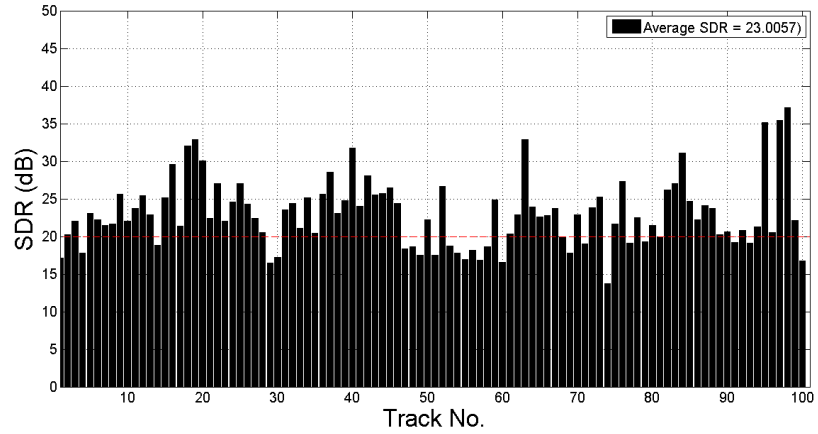

Figure A.9: SDR ($N = 816$, Fixed-parameter model, Repetitions $= 5$.)

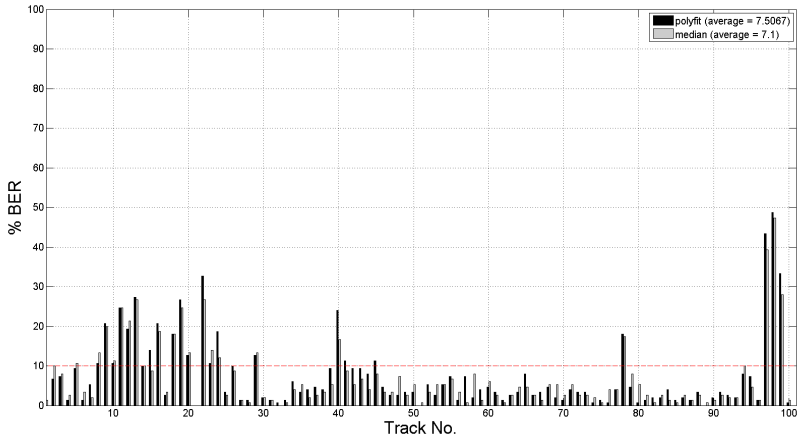

Figure A.10: BER (%) (MP3, Fixed-parameter model, $N = 2450$, No repetition.)

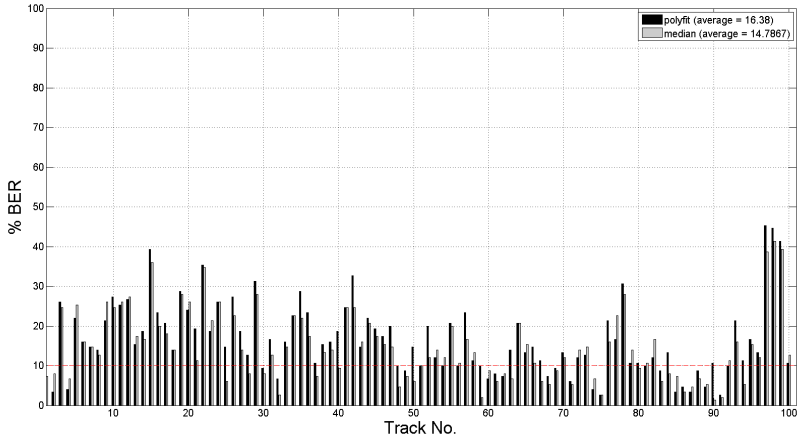

Figure A.11: BER (%) (MP3 attack, Fixed-parameter model, $N = 816$, No repetition.)

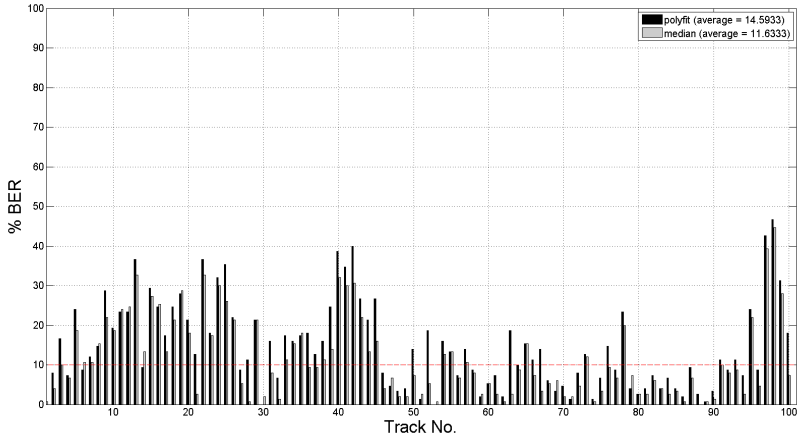

Figure A.12: BER (%) (MP3 attack, Fixed-parameter model, $N = 816$, Repetitions $= 5$.)

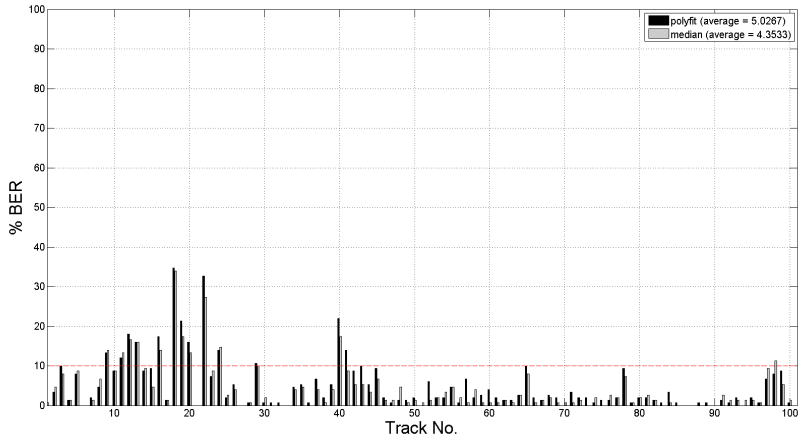

Figure A.13: BER (%) (MP4, Fixed-parameter model, $N = 2450$, No repetition.)

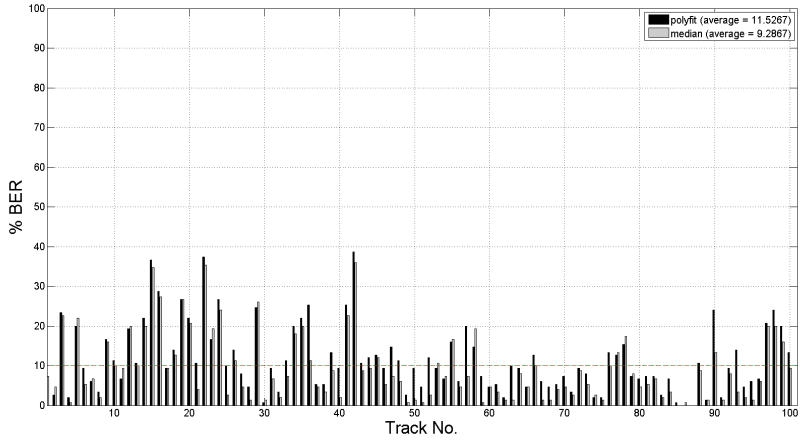

Figure A.14: BER (%) (MP4 attack, Fixed-parameter model, $N = 816$, No repetition.)

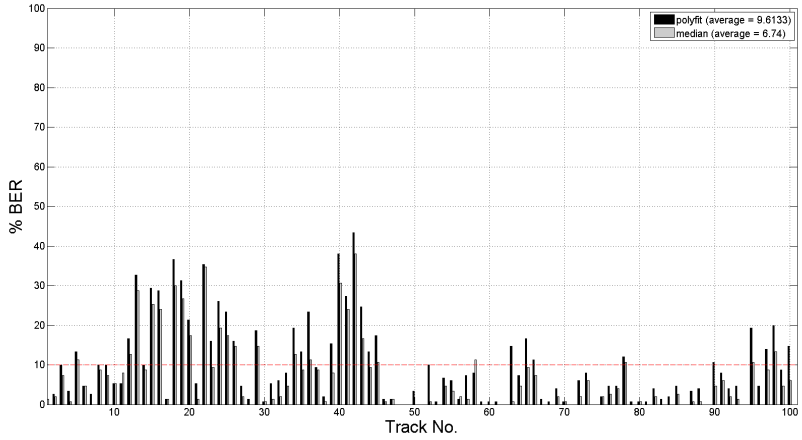

Figure A.15: BER (%) (MP4 attack, Fixed-parameter model, $N = 816$, Repetitions $= 5$.)

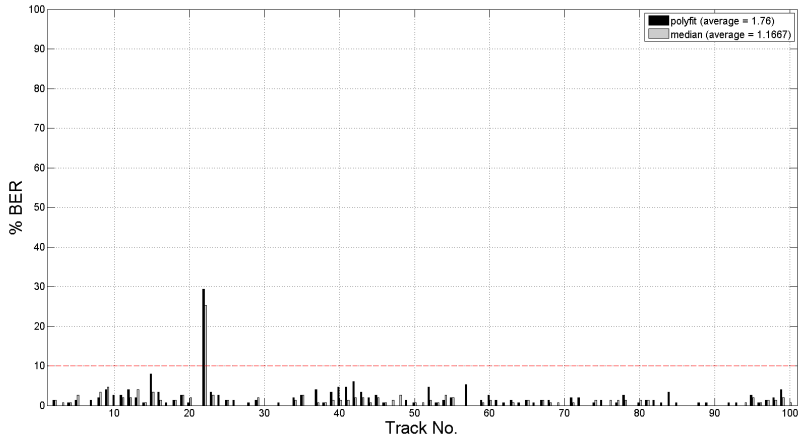

Figure A.16: BER (%) (AWGN, Fixed-parameter model, $N = 2450$, No repetition.)

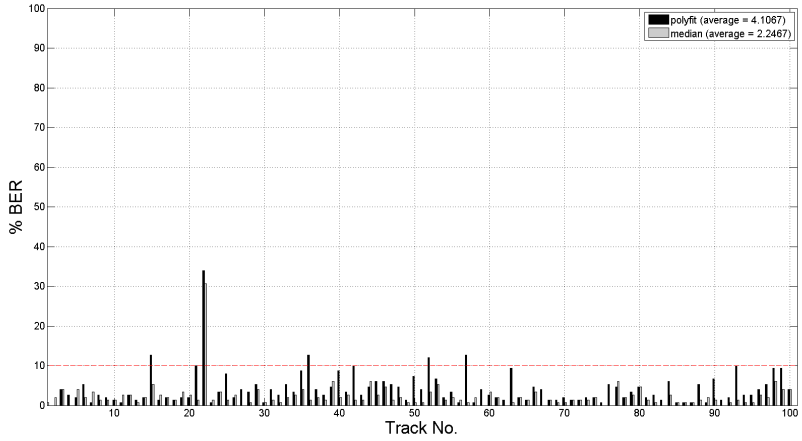

Figure A.17: BER (%) (AWGN, Fixed-parameter model, $N = 816$, No repetition.)

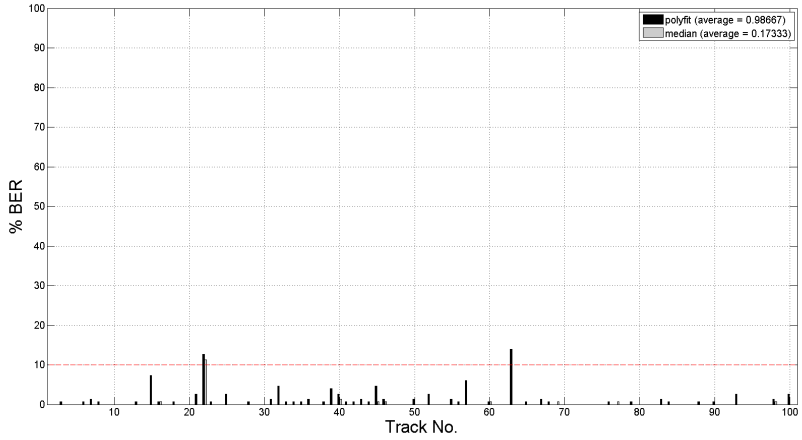

Figure A.18: BER (%) (AWGN, Fixed-parameter model, $N = 816$, Repetitions $= 5$.)

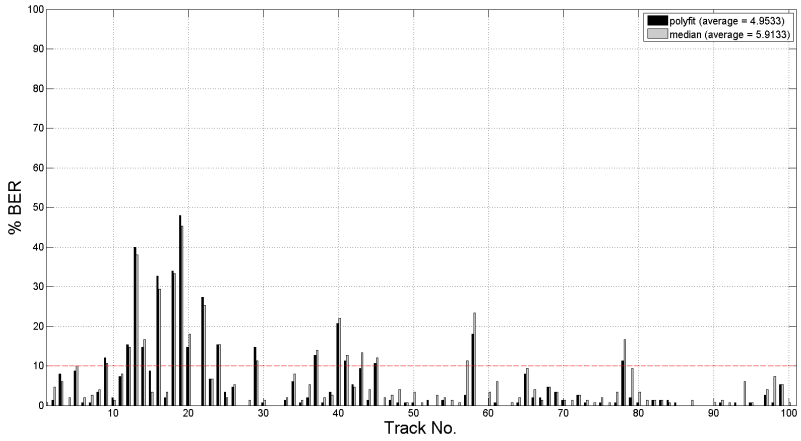

Figure A.19: BER (%) (BPF, Fixed-parameter model, $N = 2450$, No repetition.)

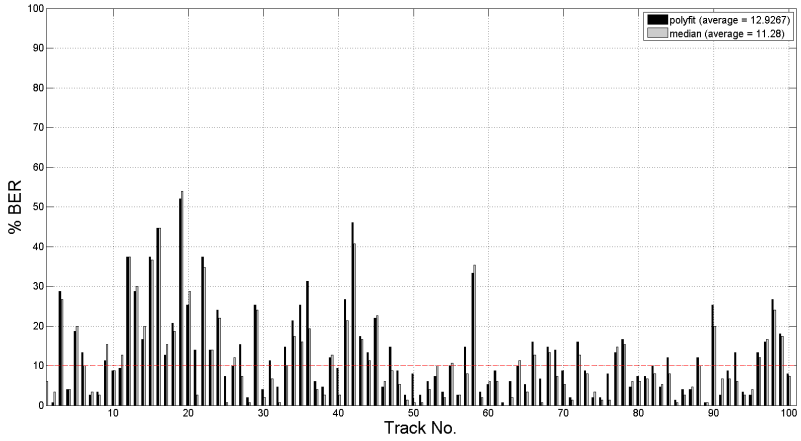

Figure A.20: BER (%) (BPF, Fixed-parameter model, $N = 816$, No repetition.)

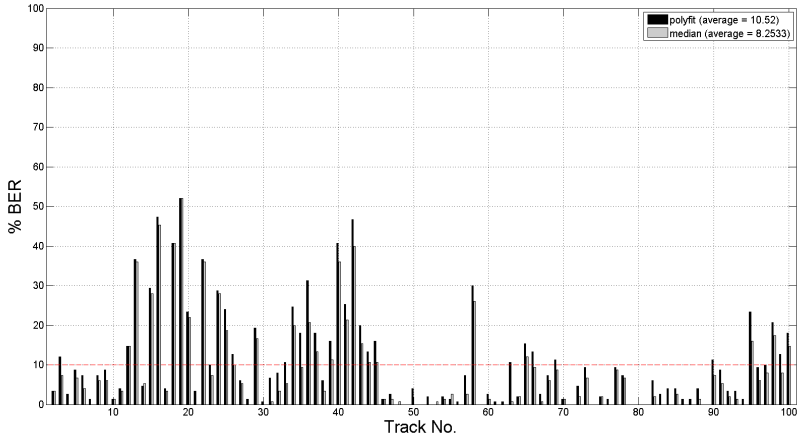

Figure A.21: BER (%) (BPF, Fixed-parameter model, $N = 816$, Repetitions $= 5$.)

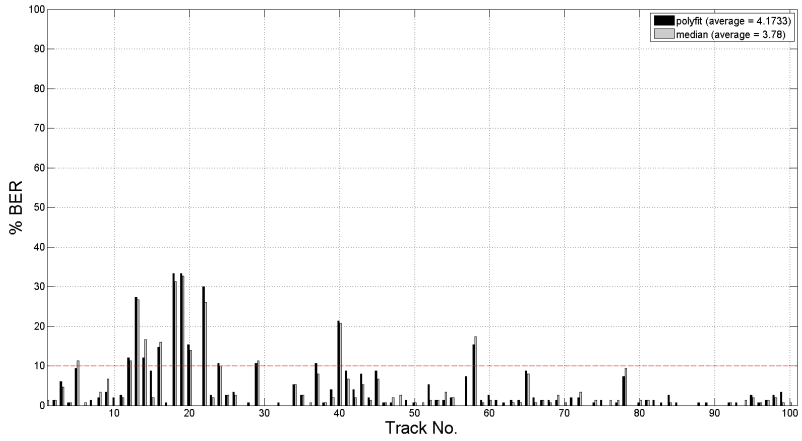

Figure A.22: BER (%) (RES 16, Fixed-parameter model, $N = 2450$, No repetition.)

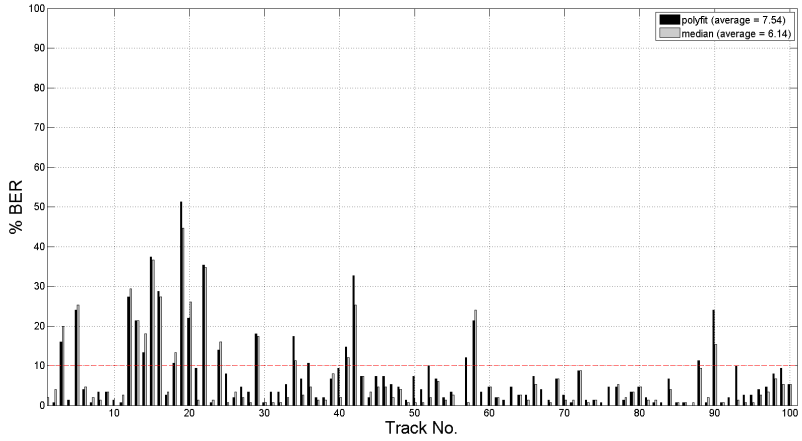

Figure A.23: BER (%) (RES 16, Fixed-parameter model, $N = 816$, No repetition.)

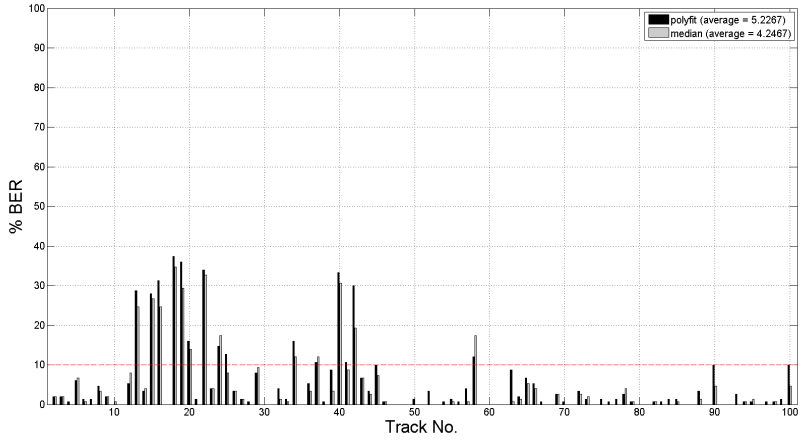

Figure A.24: BER (%) (RES 16, Fixed-parameter model, $N = 816$, Repetitions $= 5$.)

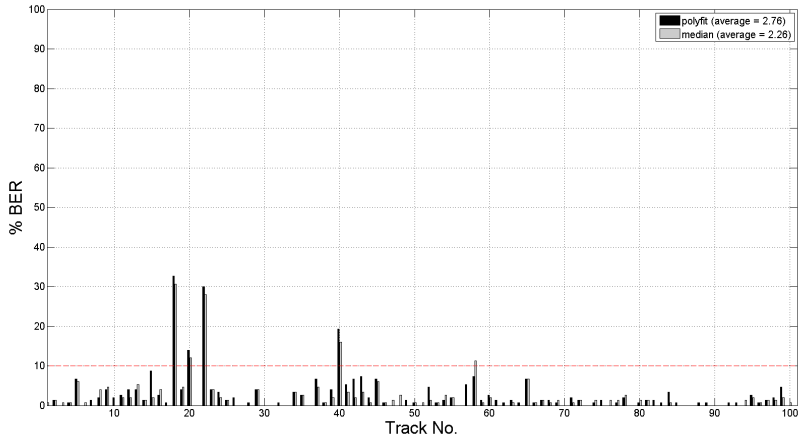

Figure A.25: BER (%) (RES 22.05, Fixed-parameter model, $N = 2450$, No repetition.)

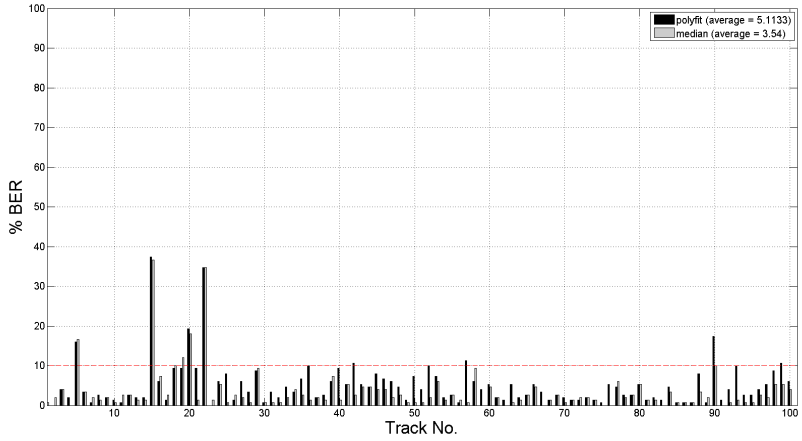

Figure A.26: BER (%) (RES 22.05, Fixed-parameter model, $N = 816$, No repetition.)

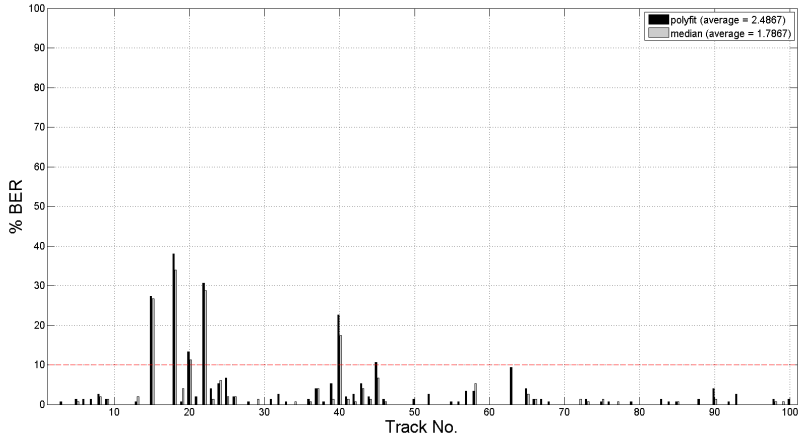

Figure A.27: BER (%) (RES 22.05, Fixed-parameter model, $N = 816$, Repetitions $= 5$.)

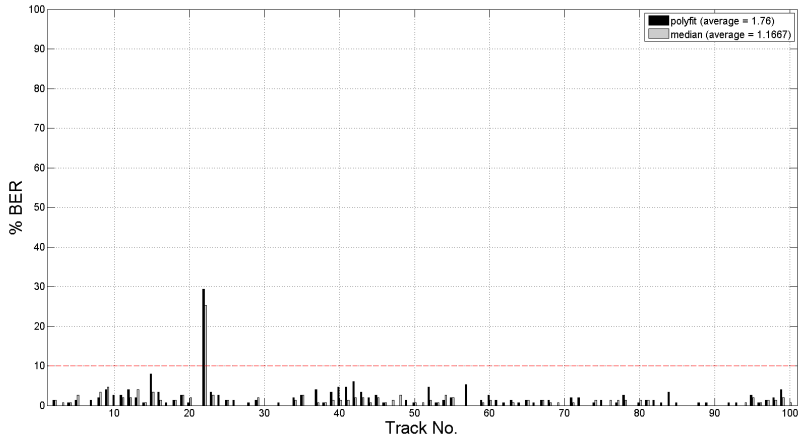

Figure A.28: BER (%) (No attack, Fixed-parameter model, $N = 2450$, No repetition.)

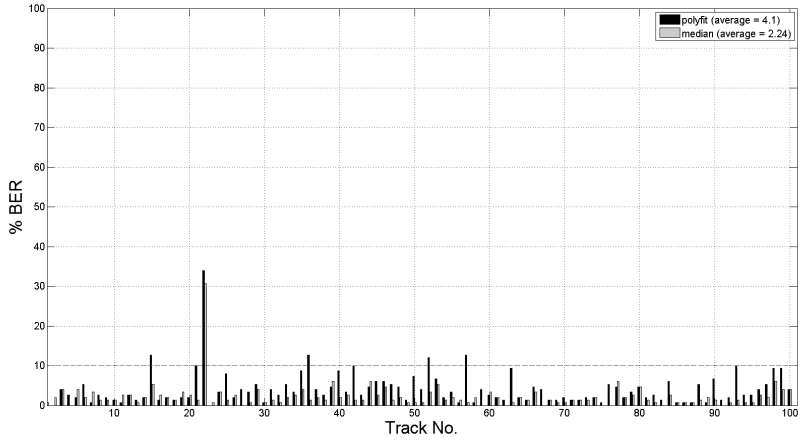

Figure A.29: BER (%) (No attack, Fixed-parameter model, $N = 816$, No repetition.)

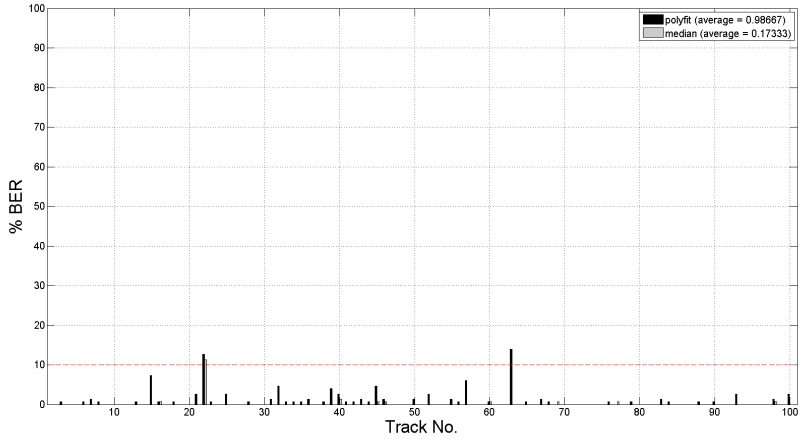

Figure A.30: BER (%) (No attack, Fixed-parameter model, $N = 816$, Repetitions $= 5$.)

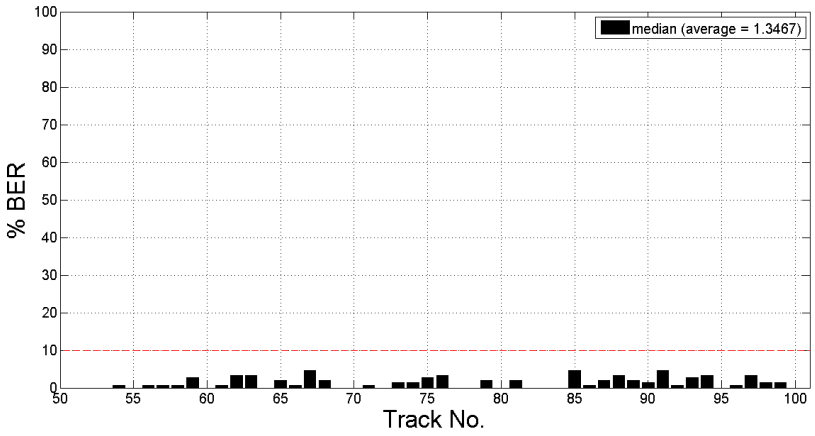

Figure A.31: BER (%) (No attack, Partially-blind model, $N = 2450$, No repetition.)

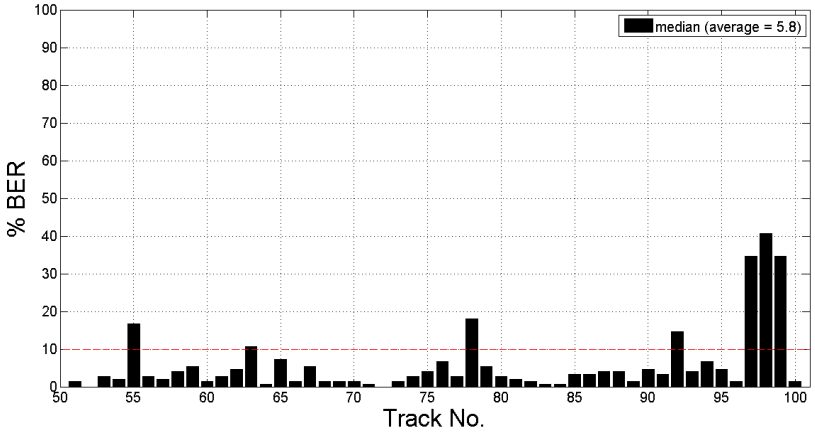

Figure A.32: BER (%) (MP3, Partially-blind model, $N = 2450$, No repetition.)

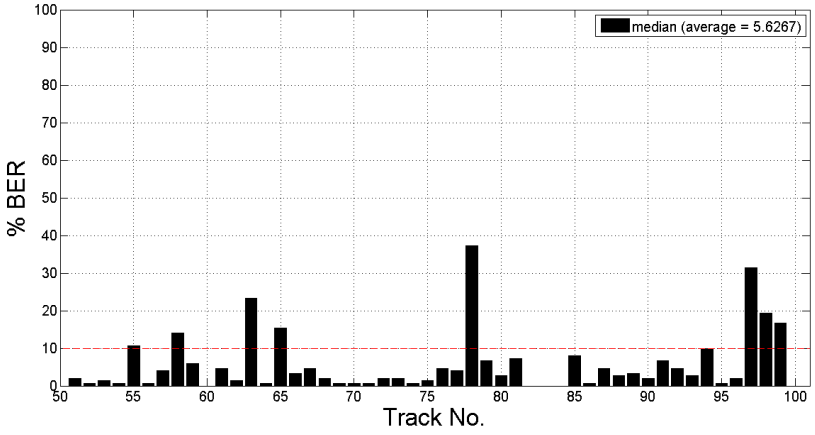

Figure A.33: BER (%) (BPF, Partially-blind model, $N = 2450$, No repetition.)

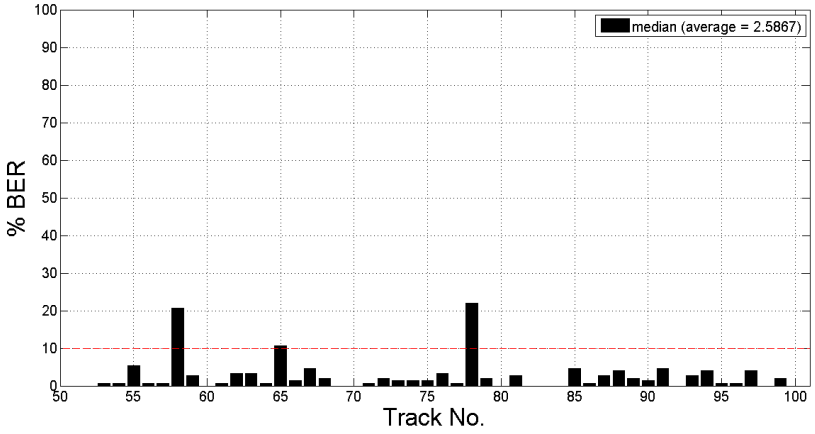

Figure A.34: BER (%) (RES 16, Partially-blind model, $N = 2450$, No repetition.)

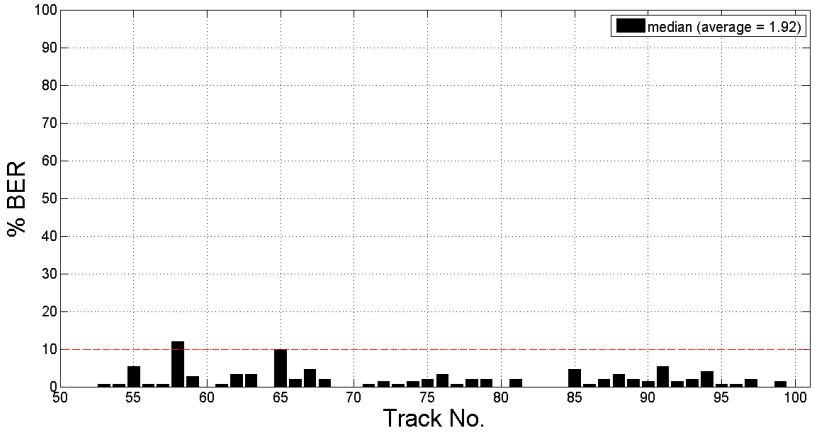

Figure A.35: BER (%) (RES 22.05, Partially-blind model, $N = 2450$, No repetition.)

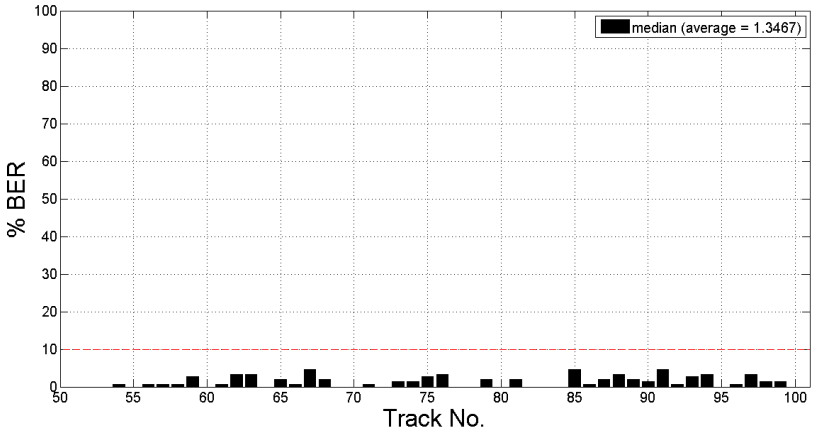

Figure A.36: BER (%) (AWGN, Partially-blind model, $N = 2450$, No repetition.)

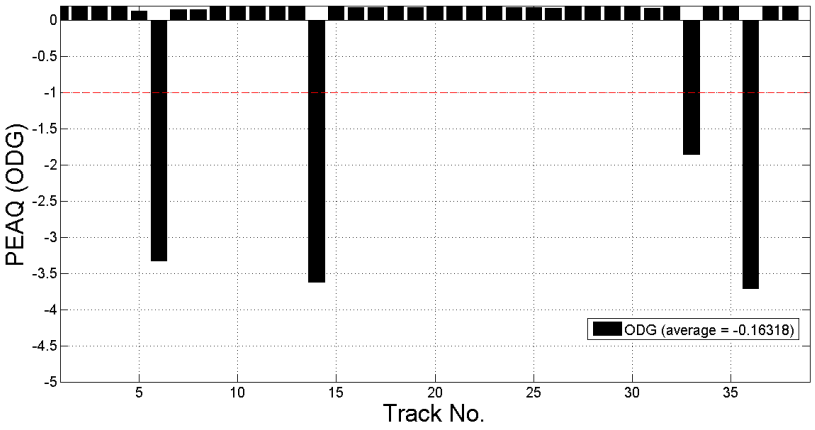

Figure A.37: PEAQ ($N = 2450$, Completely-blind model, No repetition.)

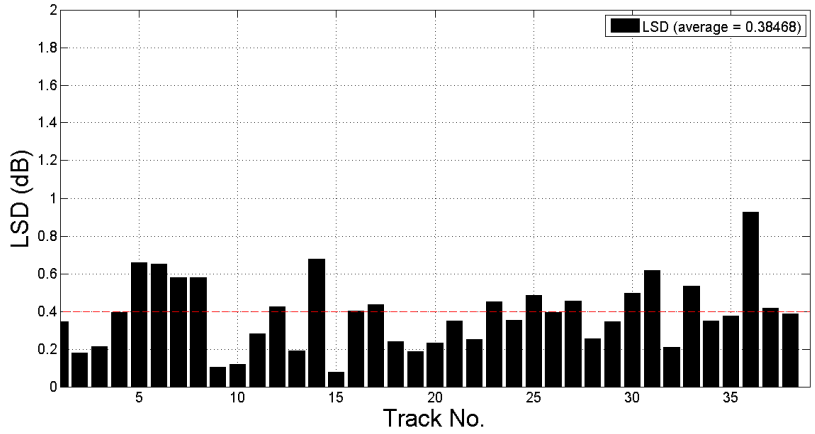

Figure A.38: LSD ($N = 2450$, Completely-blind model, No repetition.)

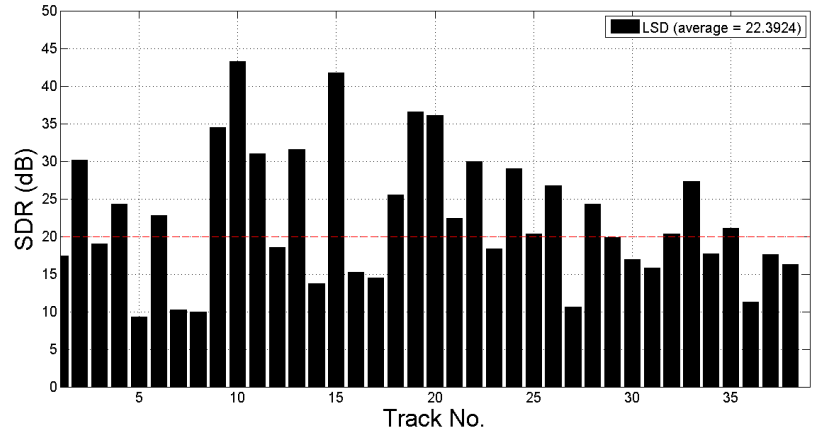

Figure A.39: SDR ($N = 2450$, Completely-blind model, No repetition.)

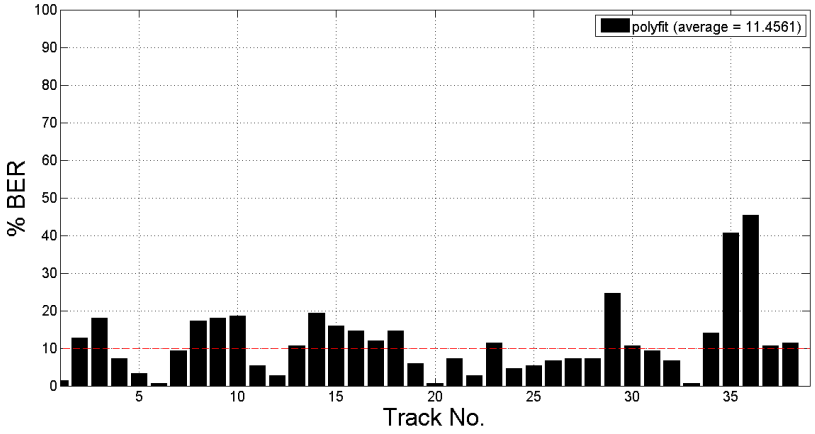

Figure A.40: BER (%) (No attack, Completely-blind model, $N = 2450$, No repetition.)

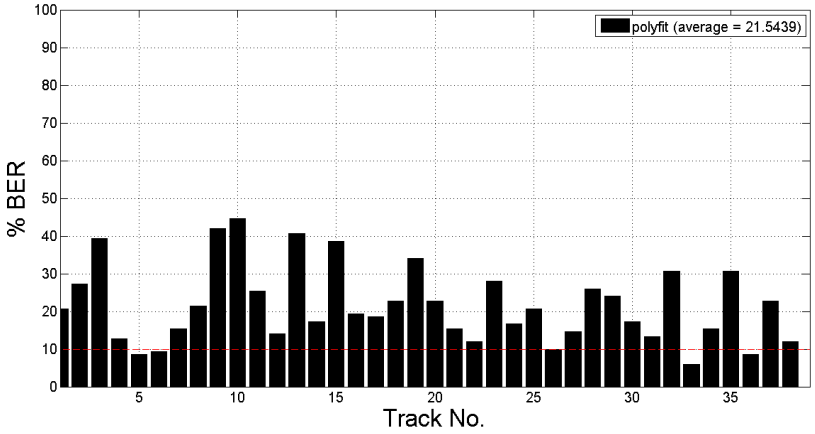

Figure A.41: BER (%) (MP3, Completely-blind model, $N = 2450$, No repetition.)

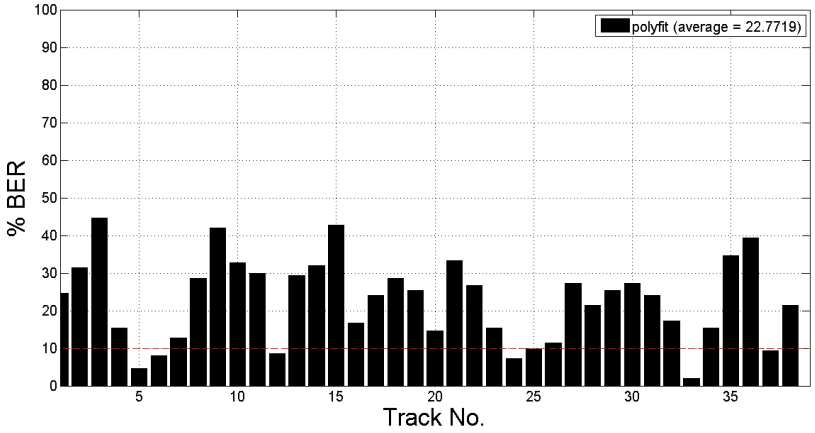

Figure A.42: BER (%) (BPF, Completely-blind model, $N = 2450$, No repetition.)

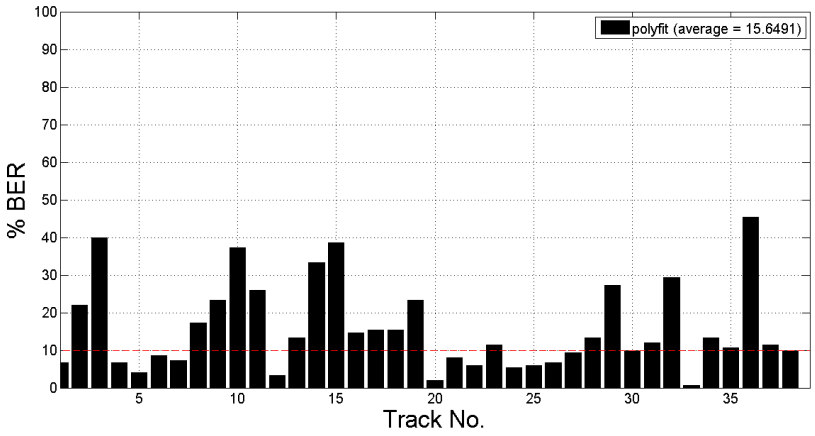

Figure A.43: BER (%) (RES 16, Completely-blind model, $N = 2450$, No repetition.)

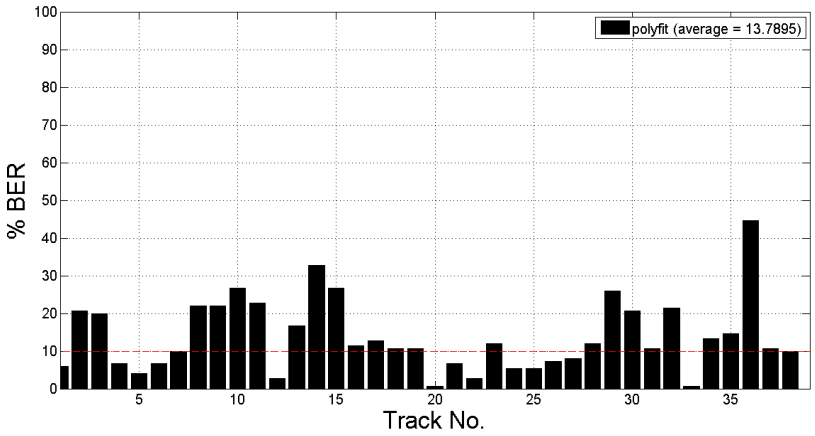

Figure A.44: BER (%) (RES 22.05, Completely-blind model, $N = 2450$, No repetition.)

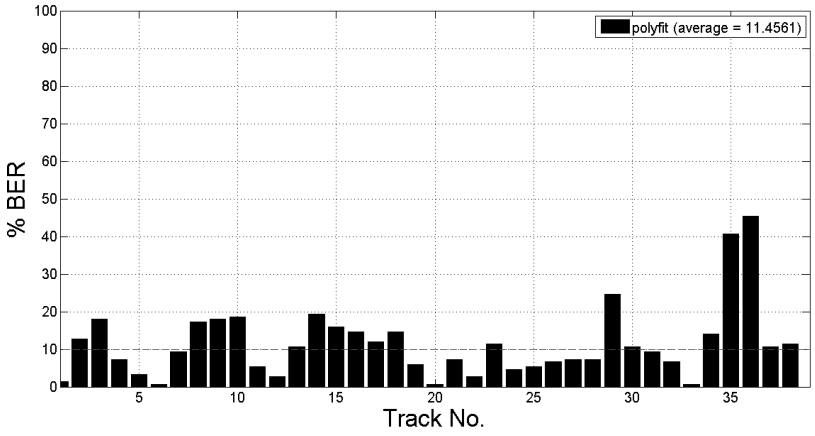

Figure A.45: BER (%) (AWGN, Completely-blind model, $N = 2450$, No repetition.)

Appendix B

Optimized parameters

Table B.1: Parameters I (Partially-blind model).

	ϵ	u	l	cost
1	0	11	43	0.83
2	0.025	33	71	0.61
3	0	10	32	4.48
4	0.075	23	55	0.63
5	0	11	51	4.72
6	0	10	54	1.41
7	0.05	17	63	2.16
8	0	20	56	1.51
9	0	109	169	5.42
10	0	10	58	4.02
11	0	90	132	0.62
12	0	94	154	3.09
13	0.025	19	49	2.84
14	0.025	10	66	1.51
15	0.175	10	34	5.7
16	0	65	119	5.19
17	0	14	72	2.91
18	0	13	67	2.17
19	0.075	10	34	7.09
20	0	13	57	5.2
21	0.025	13	33	1
22	0.075	11	37	19.57
23	0.075	105	165	3.54
24	0	10	30	4.02
25	0.05	14	50	1.47

	ϵ	u	l	cost
26	0	12	64	2.86
27	0.075	10	34	1.1
28	0.025	12	34	1.02
29	0	10	34	5.43
30	0.075	10	54	1.47
31	0.05	11	37	0.66
32	0.05	18	50	0.87
33	0.075	11	35	1.26
34	0	10	40	1.96
35	0.025	10	38	2.4
36	0.1	10	42	1.25
37	0	10	40	1.06
38	0.05	12	38	1.32
39	0	15	59	3.56
40	0.025	14	36	1.68
41	0	11	51	5.9
42	0.025	10	32	1.23
43	0.025	13	43	2.16
44	0.05	110	168	5.89
45	0	83	143	3.09
46	0.075	13	55	1.77
47	0.025	20	52	2.4
48	0.05	11	33	0.79
49	0.125	10	38	1.14
50	0.075	15	39	0.71

Table B.2: Parameters II (Partially-blind model).

	ϵ	u	l	cost
51	0.025	17	45	0.99
52	0.025	10	34	1.1
53	0	10	50	2.41
54	0.05	13	43	1.32
55	0	105	165	4.72
56	0.05	10	46	0.96
57	0.05	10	30	0.97
58	0.05	19	77	2.39
59	0.15	22	42	0.54
60	0	10	54	2.21
61	0.175	10	42	1.33
62	0.175	10	36	0.94
63	0.125	13	33	0.71
64	0	11	53	3.11
65	0.1	23	51	1.47
66	0.05	10	34	1.25
67	0.05	18	40	0.71
68	0	23	55	0.52
69	0.05	10	56	1.21
70	0	17	53	0.75
71	0.075	10	46	0.79
72	0	11	43	0.66
73	0.125	10	40	1.27
74	0	13	43	1.17
75	0	26	46	0.57

	ϵ	u	l	cost
76	0.05	13	33	0.81
77	0	35	95	3.55
78	0	72	130	4.48
79	0	10	40	1.27
80	0	13	59	3.11
81	0.1	11	49	1.49
82	0.05	19	57	2.37
83	0.075	13	45	1.25
84	0	11	43	0.89
85	0.075	11	41	1.05
86	0.1	14	38	0.81
87	0.175	10	36	1.07
88	0.075	12	40	0.77
89	0	13	41	0.99
90	0.025	18	38	0.73
91	0.075	11	43	0.82
92	0	64	122	4.01
93	0.025	11	39	1.4
94	0.1	14	52	1.56
95	0	39	87	1.67
96	0.025	11	39	0.91
97	0.025	90	146	4.95
98	0	67	125	7.07
99	0	76	134	6.13
100	0.025	10	46	1.74

Table B.3: Parameters I (Completely-blind model).

	e	u	l	cost
1	0	34	92	10.19
7	0.025	68	126	11.69
13	0	18	68	13.58
28	0	40	90	7.55
31	0	10	54	14.72
32	0.025	32	84	7.93
33	0	10	56	16.6
34	0	10	58	15.47
35	0.025	77	137	19.99
36	0.1	69	119	26.78
37	0	43	97	12.82
38	0	22	70	10.94
39	0	55	113	15.46
40	0	11	67	7.18
41	0.025	94	148	20.74
42	0	15	59	16.6
43	0	14	64	12.07
44	0	31	83	15.84
45	0.025	77	133	12.82

	e	u	l	cost
46	0	62	118	20.74
47	0	45	105	12.07
48	0	65	119	10.19
49	0.025	26	72	8.69
50	0	40	94	7.17
51	0	37	91	7.93
52	0.025	35	85	13.2
53	0	13	71	9.81
54	0	51	107	10.56
55	0.05	31	81	16.97
56	0	29	89	9.06
57	0.025	21	69	7.56
58	0	36	84	15.84
59	0.025	37	91	5.29
60	0	30	80	10.56
64	0	31	91	13.96
85	0	10	60	3.87
91	0	27	81	5.68
100	0	29	79	16.59

Table B.4: Parameters (Psychoacoustic model).

		1	**7**	**13**	**28**	**37**	**49**	**54**	**57**	**64**	**85**	**91**	**100**
Simulation 1	u	80	100	60	70	100	80	80	150	80	100	80	90
	l	150	150	100	150	200	150	150	250	150	200	150	160
Simulation 2	u	80	100	60	70	100	80	80	150	80	100	80	90
	l	150	150	100	150	200	150	150	250	150	200	150	160
Simulation 3	u	40	50	30	35	50	40	40	75	40	50	40	45
	l	100	100	70	100	130	100	100	160	100	130	100	103
Simulation 4	u	30	40	20	45	40	40	30	75	30	60	40	35
	l	90	90	60	110	120	100	90	160	90	140	100	93

Appendix C

Evaluation of the parameter estimation methods

Twelve signals from RWC music-genre database [133] (Tracks 01, 07, 13, 28, 37, 49, 54, 57, 64, 85, 91, and 100) were used. The first 294000 samples of one channel of each signal were used in our experiments. Each 294000-sample signal was divided into 120 frames, where each frame consists of 2450 samples. Singular spectrum on the interval $[u, l]$ of each frame was modified. The values of u and l were determined by the differential evolution optimization. The window length L for matrix formation in the basic SSA was set to 500.

In order to implement the concavity density-based method, we set the offsetting constant for adjusting indices at the rising and falling edges of the positive average-density curve as follows. Given σ and τ are the indices at rising and falling edges respectively, the estimated u is $\lceil \sigma + 0.3 \times (\tau - \sigma) \rceil$, and the estimated l is $\lceil \tau + 0.7 \times (\tau - \sigma) \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. Five different lengths (15, 20, 25, 30, and 35) were used to calculate the concavity density curves.

Experimental results are shown in Table C.1. The derivative-based method could not deliver estimated l when the value of l is quite high, i.e., l is located at the tail of the singular spectrum. On the other hand, the estimation by the concavity density-based method is reasonable. The root-mean-square deviations were 4.33 and 3.33 for the estimation of u and l respectively. The indices at rising and falling edges of the average positive-density curve are shown in Table C.2.

Table C.1: Actual and estimated parameters. The symbol × denotes the estimation method cannot deliver the estimated value.

		01	**07**	**13**	**28**	**37**	**49**	**54**	**57**	**64**	**85**	**91**	**100**
Actual singular spectrum modification range	u	21	17	39	64	28	70	60	20	20	40	62	61
	l	37	33	71	84	50	100	92	28	36	66	90	95
Derivative-based estimation	u	24	18	13	26	15	46	41	19	19	10	44	56
	l	37	31	×	×	×	×	×	27	34	×	×	94
Concavity density-based estimation	u	22	17	43	59	35	71	63	17	21	48	67	66
	l	37	33	72	85	50	92	98	31	40	66	88	96

Table C.2: Indices at rising (σ) and falling (τ) edges of the average positve-density curve derived from the concavity density-based method.

	01	**07**	**13**	**28**	**37**	**49**	**54**	**57**	**64**	**85**	**91**	**100**
Actual u	21	17	39	64	28	70	60	20	20	40	62	61
Rising-edge index σ	18	13	36	53	31	66	55	14	16	44	62	59
Actual l	37	33	71	84	50	100	92	28	36	66	90	95
Falling-edge index τ	29	25	57	72	42	81	80	24	30	57	77	81

Appendix D

Comparative evaluations of inaudible and robust audio-watermarking methods

Seven methods were implemented and compared: the method based on the least-significant-bit (LSB) replacement [14], the method based on the direct spread spectrum (DSS) [157], the echo hiding method (ECHO) [158], the conventional SVD-based method [23], and the three methods based on the cochlear delay characteristics, which are the non-blind [119], the blind [120], and the reversible ones [17]. Five attacks were performed on watermarked signals: Gaussian-noise addition with average signal-to-nois ratio (SNR) of 36 dB, re-sampling with 16 and 22.05 kHz, band-pass filtering with 100-6000 Hz and -12 dB/Oct, MP3 compression with 128 kbps joint stereo, and MP4 compression with 96 kbps. The average BERs are shown in Table D.1.

Table D.1: Average BER (%): comparison of the fixed-parameter model and the typical methods.

	DSS	ECHO	LSB	Non-blind	Blind	Reversible	SVD	Fixed
BER (%)	14.28	2.89	35.40	5.32	5.25	5.25	14.74	3.68

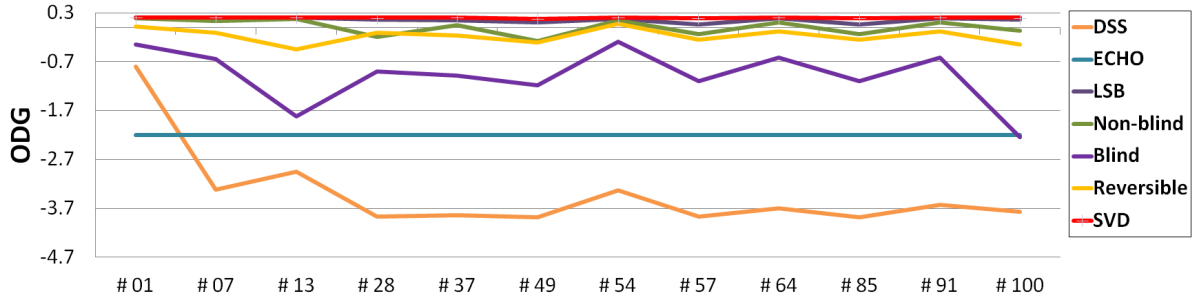

Figure D.1: PEAQ Comparison.

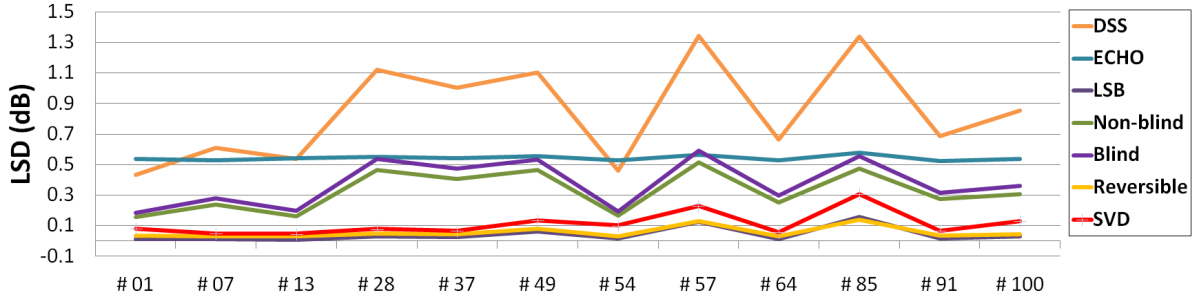

Figure D.2: LSD Comparison.

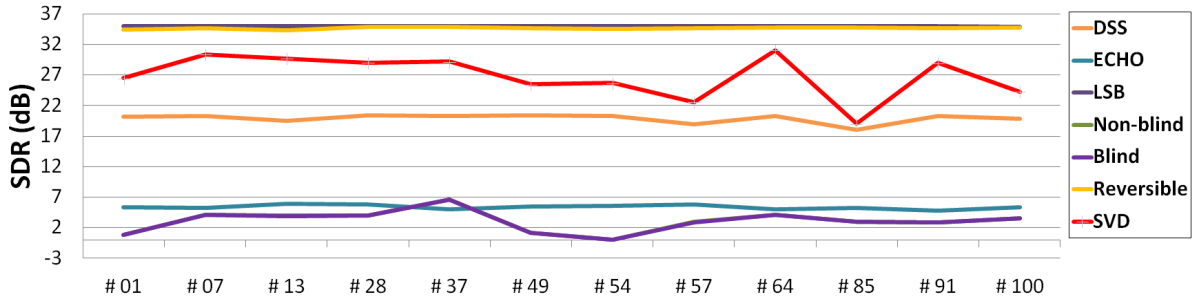

Figure D.3: SDR Comparison.

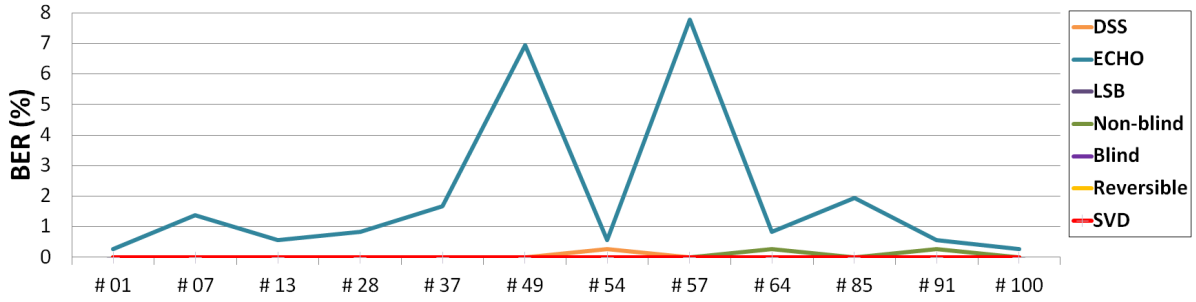

Figure D.4: BER Comparison (No attack).

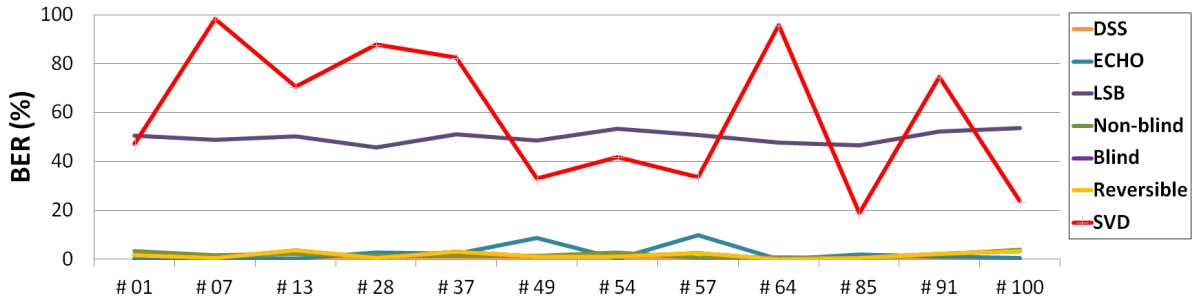

Figure D.5: BER Comparison (MP3 compression).

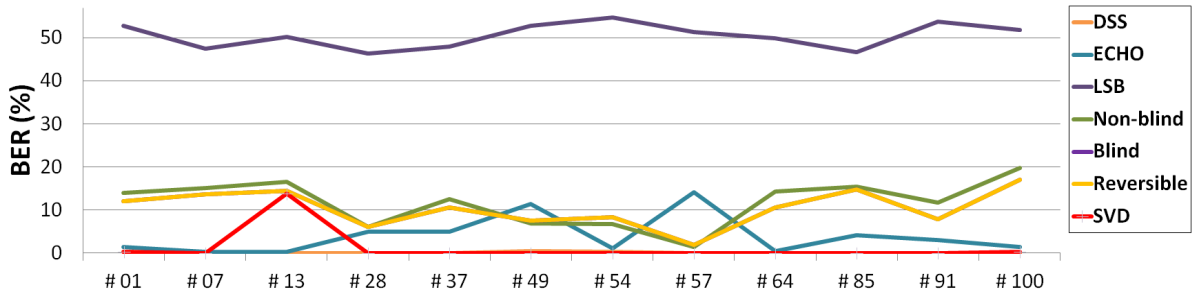

Figure D.6: BER Comparison (MP4 compression).

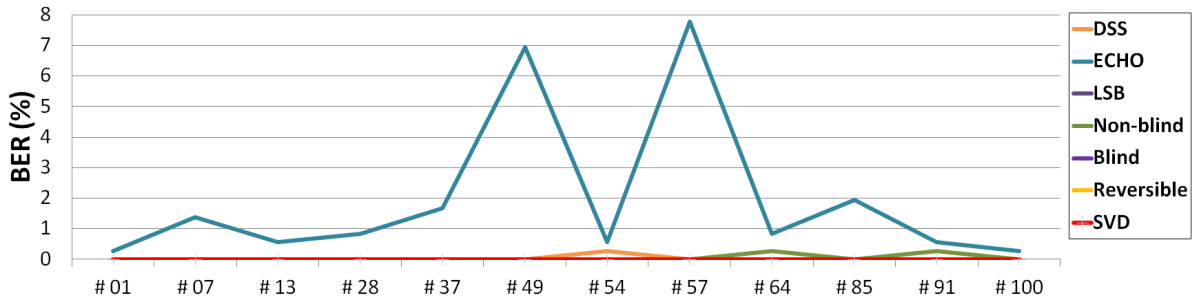

Figure D.7: BER Comparison (AWGN).

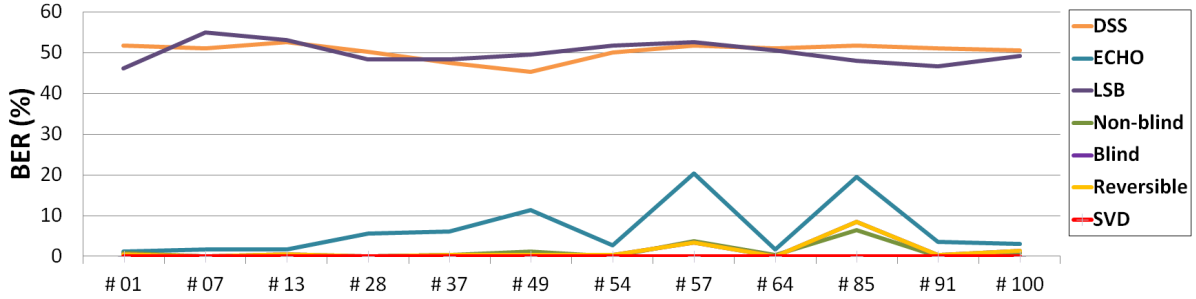

Figure D.8: BER Comparison (8-bit re-quantization).

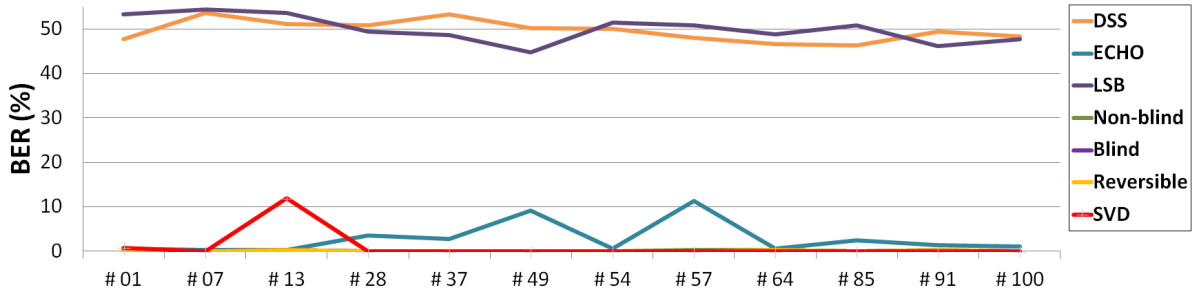

Figure D.9: BER Comparison (22.05 kHz re-sampling).

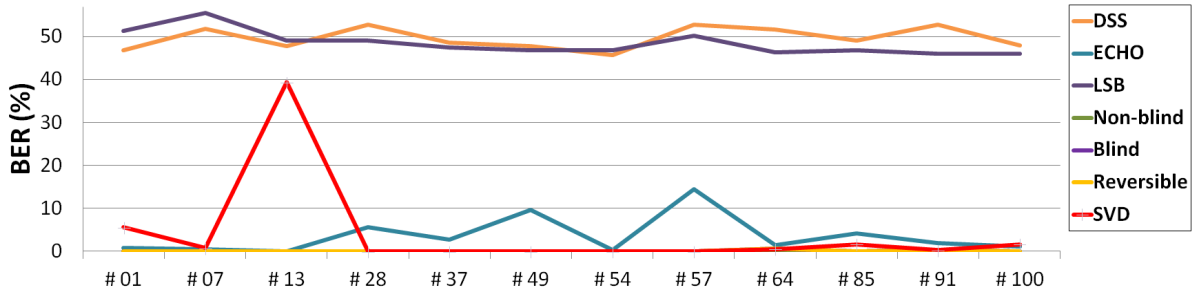

Figure D.10: BER Comparison (16 kHz re-sampling).

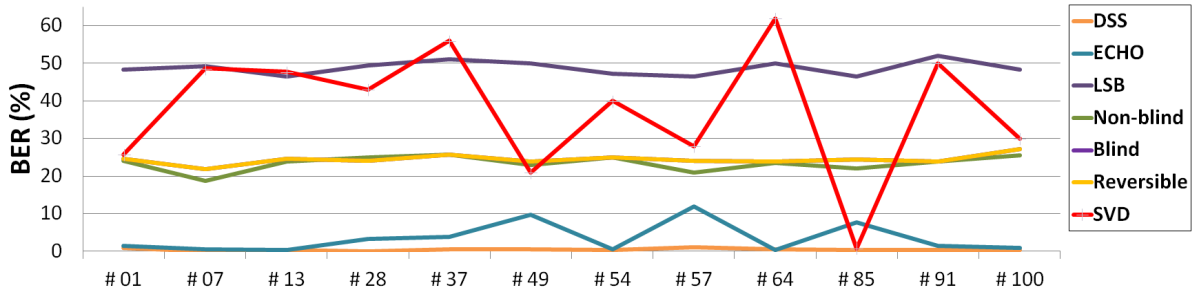

Figure D.11: BER Comparison (Band-pass filtering).

Bibliography

[1] T. Dutoit and F. Marques, *Applied Signal Processing: A MATLABTM-based proof of concept*. Springer Science & Business Media, 2010.

[2] R. D. Gopal and G. L. Sanders, "Global software piracy: You can't get blood out of a turnip," *Communications of the ACM*, vol. 43, no. 9, pp. 82–89, 2000.

[3] S. B. Weinstein, *The multimedia internet*. Springer, 2005.

[4] D. Pan, "A tutorial on mpeg/audio compression," *IEEE multimedia*, no. 2, pp. 60–74, 1995.

[5] M. Givoni, "Development and impact of the modern high-speed train: A review," *Transport reviews*, vol. 26, no. 5, pp. 593–611, 2006.

[6] T. C. Kwong and M. K. Lee, "Behavioral intention model for the exchange mode internet music piracy," in *System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on*, pp. 2481–2490, IEEE, 2002.

[7] S. Bhattacharjee, R. D. Gopal, and G. L. Sanders, "Digital music and online sharing: software piracy 2.0?," *Communications of the ACM*, vol. 46, no. 7, pp. 107–111, 2003.

[8] I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker, *Digital watermarking and steganography*. Morgan Kaufmann, 2007.

[9] A. M. Eskicioglu and E. J. Delp, "An overview of multimedia content protection in consumer electronics devices," *Signal Processing: Image Communication*, vol. 16, no. 7, pp. 681–699, 2001.

[10] A. J. Menezes, P. C. Van Oorschot, and S. A. Vanstone, *Handbook of applied cryptography*. CRC press, 1996.

[11] M. Peitz and P. Waelbroeck, "Why the music industry may gain from free downloadingthe role of sampling," *International Journal of Industrial Organization*, vol. 24, no. 5, pp. 907–913, 2006.

[12] T. Sharp, "An implementation of key-based digital signal steganography," in *Information hiding*, pp. 13–26, Springer, 2001.

[13] A. M. Al-Haj, *Advanced techniques in multimedia watermarking: image, video and audio applications: image, video and audio applications*. IGI Global, 2010.

[14] N. Cvejic, *Digital Audio Watermarking Techniques and Technologies: Applications and Benchmarks: Applications and Benchmarks*. IGI Global, 2007.

[15] S. Craver, M. Wu, B. Liu, A. Stubblefield, B. Swartzlander, D. S. Wallach, D. Dean, and E. W. Felten, "Reading between the lines: Lessons from the sdmi challenge.," in *USENIX Security Symposium*, 2001.

[16] S. A. Craver, M. Wu, and B. Liu, "What can we reasonably expect from watermarks?," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pp. 223–226, IEEE, 2001.

[17] M. Unoki and R. Miyauchi, "Reversible watermarking for digital audio based on cochlear delay characteristics," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2011 Seventh International Conference on*, pp. 314–317, IEEE, 2011.

[18] A. Al-Haj, C. Twal, and A. Mohammad, "Hybrid dwt-svd audio watermarking," in *Digital Information Management (ICDIM), 2010 Fifth International Conference on*, pp. 525–529, IEEE, 2010.

[19] B. Lei, Y. Soon, and E.-L. Tan, "Robust svd-based audio watermarking scheme with differential evolution optimization," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 21, no. 11, pp. 2368–2378, 2013.

[20] M. Fallahpour and D. Megias, "High capacity audio watermarking using fft amplitude interpolation," *IEICE Electronics Express*, vol. 6, no. 14, pp. 1057–1063, 2009.

[21] C. J. Plack, *The sense of hearing*. Psychology Press, 2013.

[22] Buried inside a Bruce Willis video, the evidence of a plot to kill thousands: `http://www.theguardian.com/uk/2006/nov/07/terrorism.world` [Accessed 22 May 2016].

[23] V. Bhat, I. Sengupta, and A. Das, "A new audio watermarking scheme based on singular value decomposition and quantization," *Circuits, Systems, and Signal Processing*, vol. 30, no. 5, pp. 915–927, 2011.

[24] L. Lamarche, Y. Liu, and J. Zhao, "Flaw in svd-based watermarking," in *Electrical and Computer Engineering, 2006. CCECE'06. Canadian Conference on*, pp. 2082–2085, IEEE, 2006.

[25] H. Özer, B. Sankur, and N. Memon, "An svd-based audio watermarking technique," in *Proc. The Workshop on Multimedia and Security*, (New York), pp. 51–56, 2005.

[26] V. Bhat, I. Sengupta, and A. Das, "Audio watermarking based on quantization in wavelet domain," in *Information Systems Security*, pp. 235–242, Springer, 2008.

[27] V. Bhat, I. Sengupta, and A. Das, "An audio watermarking scheme using singular value decomposition and dither-modulation quantization," *Multimedia Tools and Applications*, vol. 52, no. 2-3, pp. 369–383, 2011.

[28] F. E. A. El-Samie, "An efficient singular value decomposition algorithm for digital audio watermarking," *International Journal of Speech Technology*, vol. 12, no. 1, pp. 27–45, 2009.

[29] S. Vongpraphip and M. Ketcham, "An intelligence audio watermarking based on dwt-svd using ats," in *Intelligent Systems, 2009. GCIS'09. WRI Global Congress on*, vol. 3, pp. 150–154, IEEE, 2009.

[30] A. Al-Haj and A. Mohammad, "Digital audio watermarking based on the discrete wavelets transform and singular value decomposition," *European Journal of Scientific Research*, vol. 39, no. 1, pp. 6–21, 2010.

[31] M. S. Al-Yaman, M. A. Al-Taee, and H. A. Alshammas, "Audio-watermarking based ownership verification system using enhanced dwt-svd technique," in *Systems, Signals and Devices (SSD), 2012 9th International Multi-Conference on*, pp. 1–5, IEEE, 2012.

[32] A. Singhal, A. N. Chaubey, and C. Prakash, "Audio watermarking using combination of multilevel wavelet decomposition, dct and svd," in *Emerging Trends in Networks and Computer Communications (ETNCC), 2011 International Conference on*, pp. 239–243, IEEE, 2011.

[33] P. K. Dhar and T. Shimamura, "Audio watermarking in transform domain based on singular value decomposition and quantization," in *Communications (APCC), 2012 18th Asia-Pacific Conference on*, pp. 516–521, IEEE, 2012.

[34] P. K. Dhar and T. Shimamura, "An svd-based audio watermarking using variable embedding strength and exponential-log operations," in *Informatics, Electronics & Vision (ICIEV), 2013 International Conference on*, pp. 1–6, IEEE, 2013.

[35] P. K. Dhar and T. Shimamura, "An audio watermarking scheme using discrete fourier transformation and singular value decomposition," in *Telecommunications and Signal Processing (TSP), 2012 35th International Conference on*, pp. 789–794, IEEE, 2012.

[36] P. K. Dhar and T. Shimamura, "Audio watermarking in transform domain based on singular value decomposition and cartesian-polar transformation," *International Journal of Speech Technology*, vol. 17, no. 2, pp. 133–144, 2014.

[37] G. Suresh, N. Lalitha, C. S. Rao, and V. Sailaja, "An efficient and simple audio watermarking using dct-svd," in *Devices, Circuits and Systems (ICDCS), 2012 International Conference on*, pp. 177–181, IEEE, 2012.

[38] S. Karimimehr, S. Samavi, H. R. Kaviani, and M. Mahdavi, "Robust audio watermarking based on hwd and svd," in *Electrical Engineering (ICEE), 2012 20th Iranian Conference on*, pp. 1363–1367, IEEE, 2012.

[39] R. Zezula *et al.*, "Audio digital watermarking algorithm based on svd in mclt domain," in *Third International Conference on Systems*, pp. 140–143, IEEE, 2008.

[40] A. Verma, A. Murarka, A. Vashist, and M. K. Dutta, "An efficient audio watermarking scheme based on svd in wavelet domain and chaotic mapping," in *Engineering (NUiCONE), 2012 Nirma University International Conference on*, pp. 1–5, IEEE, 2012.

[41] M. El-Bendary, A. Haggag, F. Shawki, and F. Abd-El-Samie, "Proposed approach for improving bluetooth networks security through svd audio watermarking," in *Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2012 6th International Conference on*, pp. 594–598, IEEE, 2012.

[42] R. Sobha and M. Sucharitha, "Secure transmission of data using audio watermarking with protection on synchronization attack," in *Communication Technologies (GCCT), 2015 Global Conference on*, pp. 592–597, IEEE, 2015.

[43] M. A. Nematollahi, S. A. R. Al-Haddad, S. Doraisamy, and M. Saripan, "Digital audio and speech watermarking based on the multiple discrete wavelets transform and singular value decomposition," in *Modelling Symposium (AMS), 2012 Sixth Asia*, pp. 109–114, IEEE, 2012.

[44] N. F. Johnson and S. Jajodia, "Exploring steganography: Seeing the unseen," *Computer*, vol. 31, no. 2, pp. 26–34, 1998.

[45] S.-J. Lee and S.-H. Jung, "A survey of watermarking techniques applied to multimedia," in *Industrial Electronics, 2001. Proceedings. ISIE 2001. IEEE International Symposium on*, vol. 1, pp. 272–277, IEEE, 2001.

[46] D. Kirovski and H. S. Malvar, "Spread-spectrum watermarking of audio signals," *Signal Processing, IEEE Transactions on*, vol. 51, no. 4, pp. 1020–1033, 2003.

[47] A. Z. Tirkel, G. Rankin, R. Van Schyndel, W. Ho, N. Mee, and C. F. Osborne, "Electronic watermark," *Digital Image Computing, Technology and Applications (DICTA93)*, pp. 666–673, 1993.

[48] L. d. C. Gomes, P. Cano, E. Gomez, M. Bonnet, and E. Batlle, "Audio watermarking and fingerprinting: for which applications?," *Journal of New Music Research*, vol. 32, no. 1, pp. 65–81, 2003.

[49] C. I. Podilchuk and E. J. Delp, "Digital watermarking: algorithms and applications," *Signal Processing Magazine, IEEE*, vol. 18, no. 4, pp. 33–46, 2001.

[50] Q. X.-m. Z. Hong-bin, "Fragile audio watermarking algorithm for telltale tamper proofing and authentication [j]," *Journal of Electronics and Information Technology*, vol. 8, p. 003, 2005.

[51] X. Quan and H. Zhang, "Perceptual criterion based fragile audio watermarking using adaptive wavelet packets," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 2, pp. 867–870, IEEE, 2004.

[52] G. Depovere, T. Kalker, J. Haitsma, M. Maes, L. De Strycker, P. Termont, J. Vandewege, A. Langell, C. Alm, P. Norman, *et al.*, "The viva project: digital watermarking for broadcast monitoring," in *Image Processing, 1999. ICIP 99. Proceedings. 1999 International Conference on*, vol. 2, pp. 202–205, IEEE, 1999.

[53] T. Kalker, G. Depovere, J. Haitsma, and M. J. Maes, "Video watermarking system for broadcast monitoring," in *Electronic Imaging'99*, pp. 103–112, International Society for Optics and Photonics, 1999.

[54] T. Kalker and J. Haitsma, "Efficient detection of a spatial spread-spectrum watermark in mpeg video streams," in *Image Processing, 2000. Proceedings. 2000 International Conference on*, vol. 1, pp. 434–437, IEEE, 2000.

[55] P. Cano, E. Batlle, T. Kalker, and J. Haitsma, "A review of audio fingerprinting," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 41, no. 3, pp. 271–284, 2005.

[56] P. Cano, E. Batlle, E. Gómez, L. de CT Gomes, and M. Bonnet, "Audio fingerprinting: concepts and applications," in *Computational intelligence for modelling and prediction*, pp. 233–245, Springer, 2005.

[57] J. A. Bloom, I. J. Cox, T. Kalker, J.-P. M. Linnartz, M. L. Miller, and C. B. S. Traw, "Copy protection for dvd video," *Proceedings of the IEEE*, vol. 87, no. 7, pp. 1267–1276, 1999.

[58] S. Matsumoto and M. Furukawa, "Digital copy control method, digital recording medium, digital recording medium producing apparatus, digital reproducing apparatus and digital recording apparatus," Nov. 20 2001. US Patent 6,320,829.

[59] S. Craver and J. P. Stern, "Lessons learned from sdmi," in *Multimedia Signal Processing, 2001 IEEE Fourth Workshop on*, pp. 213–218, IEEE, 2001.

[60] I. J. Cox and M. L. Miller, "The first 50 years of electronic watermarking," *EURASIP Journal on Advances in Signal Processing*, vol. 2002, no. 2, pp. 1–7, 2002.

[61] I. J. Cox, "Watermarking, steganography and content forensics.," in *SECRYPT*, pp. 29–29, 2008.

[62] N. Jenkins and J. E. Martina, "Steganography in audio," *Avaliable: http://petitcolas. net/fabien/steganography/mp3stego*, pp. 269–78, 2009.

[63] E. Erçelebi and L. Batakçı, "Audio watermarking scheme based on embedding strategy in low frequency components with a binary image," *Digital Signal Processing*, vol. 19, no. 2, pp. 265–277, 2009.

[64] W.-N. Lie and L.-C. Chang, "Robust and high-quality time-domain audio watermarking based on low-frequency amplitude modification," *Multimedia, IEEE Transactions on*, vol. 8, no. 1, pp. 46–59, 2006.

[65] S.-K. Lee and Y.-S. Ho, "Digital audio watermarking in the cepstrum domain," *Consumer Electronics, IEEE Transactions on*, vol. 46, no. 3, pp. 744–750, 2000.

[66] M. Unoki and R. Miyauchi, "Method of digital-audio watermarking based on cochlear delay characteristics," *Multimedia Information Hiding Technologies and Methodologies for Controlling Data*, pp. 42–70, 2012.

[67] T.-y. Ye, Z.-f. Ma, X.-x. Niu, and Y.-x. Yang, "A zero-watermark technology with strong robustness," *Journal of Beijing University of Posts and Telecommunications*, vol. 33, no. 3, pp. 126–129, 2010.

[68] I.-K. Yeo and H. J. Kim, "Modified patchwork algorithm: A novel audio watermarking scheme," *Speech and Audio Processing, IEEE Transactions on*, vol. 11, no. 4, pp. 381–386, 2003.

[69] N. Cvejic and T. Seppänen, "Increasing robustness of lsb audio steganography by reduced distortion lsb coding.," *J. UCS*, vol. 11, no. 1, pp. 56–65, 2005.

[70] N. Cvejic and T. Seppänen, "A wavelet domain lsb insertion algorithm for high capacity audio steganography," in *Digital Signal Processing Workshop, 2002 and the 2nd Signal Processing Education Workshop. Proceedings of 2002 IEEE 10th*, pp. 53–55, IEEE, 2002.

[71] N. Cvejic and T. Seppänen, "Increasing robustness of lsb audio steganography using a novel embedding method," in *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on*, vol. 2, pp. 533–537, IEEE, 2004.

[72] A. Mane, G. Galshetwar, and A. Jeyakumar, "Data hiding technique: Audio steganographyusing lsb technique," *International Journal of Engineering Research and Applications (IJERA)*, vol. 2, no. 3, pp. 1123–1125, 2012.

[73] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM systems journal*, vol. 35, no. 3.4, pp. 313–336, 1996.

[74] A. D. Brown, G. C. Stecker, and D. J. Tollin, "The precedence effect in sound localization," *Journal of the Association for Research in Otolaryngology*, vol. 16, no. 1, pp. 1–28, 2015.

[75] H. O. Oh, J. W. Seok, J. W. Hong, and D. H. Youn, "New echo embedding technique for robust and imperceptible audio watermarking," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, vol. 3, pp. 1341–1344, IEEE, 2001.

[76] H. J. Kim and Y. H. Choi, "A novel echo-hiding scheme with backward and forward kernels," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 8, pp. 885–889, 2003.

[77] X. Cao and L. Zhang, "Researches on echo kernels of audio digital watermarking technology based on echo hiding," in *Wireless Communications and Signal Processing (WCSP), 2011 International Conference on*, pp. 1–5, IEEE, 2011.

[78] S. Mitra and S. Manoharan, "Experiments with and enhancements to echo hiding," in *2009 Fourth International Conference on Systems and Networks Communications*, pp. 119–124, IEEE, 2009.

[79] D. V. S. Chandra, "Digital image watermarking using singular value decomposition," in *Proc. The Midwest Symposium on Circuits and Systems*, pp. 264–267, 2002.

[80] R. Liu and T. Tan, "An svd-based watermarking scheme for protecting rightful ownership," *IEEE Transactions on Multimedia*, vol. 4, pp. 121–128, Mar. 2002.

[81] V. I. Gorodetski, L. J. Popyack, V. Samoilov, and V. A. Skormin, "Svd-based approach to transparent embedding data into digital image," in *Proc. The International Workshop on Information Assurance in Computer Networks: Methods, Models, and Architectures for Network Security*, (St. Petersburg), pp. 263–274, 2001.

[82] J. Wei-zhen, "Notice of retraction fragile audio watermarking algorithm based on svd and dwt," in *Intelligent Computing and Integrated Systems (ICISS), 2010 International Conference on*, pp. 83–86, IEEE, 2010.

[83] J. Zhang, "Analysis on audio watermarking algorithm based on svd," in *Computer Science and Network Technology (ICCSNT), 2012 2nd International Conference on*, pp. 1986–1989, IEEE, 2012.

[84] X.-P. Zhang and K. Li, "Comments on" an svd-based watermarking scheme for protecting rightful ownership"," *Multimedia, IEEE Transactions on*, vol. 7, no. 3, pp. 593–594, 2005.

[85] R. Rykaczewski, "Comments on an svd-based watermarking scheme for protecting rightful ownership," *IEEE Transactions on Multimedia*, vol. 2, no. 9, pp. 421–423, 2007.

[86] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and related techniques*. Chapman and Hall/CRC, 2001.

[87] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics form experimental data," *Physica D*, vol. 20, pp. 217–236, June 1986.

[88] H. Hassani, "Singular spectrum analysis: Methodology and comparison," *Journal of Data Science*, vol. 5, pp. 239–257, Apr. 2007.

[89] S. Enshaeifar, S. Kouchaki, C. C. Took, and S. Sanei, "Quanternion singular spectrum analysis of electroencephalogram with application in sleep analysis," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 24, pp. 57–67, Jan. 2015.

[90] T. Zeng, J. Ma, and M. Dong, "Environmental noise elimination of heart sound based on singular spectrum analysis," in *Proc. Cairo International Biomedical Engineering Conference*, (Giza), pp. 158–161, 2014.

[91] G. Ghaderi, H. R. Mohseni, and S. Sanei, "Localizing heart sounds in respiratory signals using singular spectrum analysis," *IEEE Transactions on Biomedical Engineering*, vol. 58, pp. 3360–3367, Dec. 2011.

[92] R. Storn and K. Price, "Differential evolution – a simple and efficient heuristic for global optimization over continuous speces," *Journal of Global Optimization*, vol. 11, pp. 341–359, Dec. 1997.

[93] R. Lewis, V. Torczon, and M. Trosset, "Direct search method: then and now," *Journal of Computation and Applied Mathematics*, vol. 104, pp. 191–207, Dec. 2000.

[94] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, pp. 65–85, June 1994.

[95] D. Poole and A. Mackworth, *Artificial Intelligence: Foundations of Computational Agents*. Cambridge University Press, 2010.

[96] K. Price, R. M. Storn, and J. A. Lampinen, *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.

[97] C. Sammut and G. Webb, eds., *Encyclopedia of Machine Learning*, sec. Curse of Dimensionality, pp. 257–258. Springer, 2010.

[98] J. Nelder and R. Mead, "A simplex method for function minimization," *Computer Journal*, vol. 7, no. 4, pp. 308–313, 1965.

[99] W. Price, "A controlled random search procedure for global optimization," *Computer Journal*, vol. 20, no. 4, pp. 367–370, 1977.

[100] Y. Ao and H. Chi, "Experimental study on differential evolution strategies," in *Proc. WRI Global Congress on Intelligence Systems*, (Xiamen), pp. 19–24, 2009.

[101] C. Sun, H. Zhou, and L. Chen, "Improved differential evolution algorithms," in *Proc. IEEE International Conference on Computer Science and Automation Engineering*, (Zhangjiajie), pp. 142–145, 2012.

[102] G. Li and M. Liu, "The summary of differential evolution algorithm and its improvements," in *Proc. International Conference on Advanced Computer Theory and Engineering*, (Chengdu), pp. 153–156, 2010.

[103] K. Yasuda, K. Makise, and K. Tamura, "A study on combination of differential evolution and evolution strategy," in *Proc. IEEE International Conference on Systems, Man, and Cybernetics*, (Seoul), pp. 587–592, 2012.

[104] Y. Lin and W. H. Abdulla, *Audio Watermark*. Springer, 2015.

[105] A. Spanias, T. Painter, and V. Atti, *Audio signal processing and coding*. John Wiley & Sons, 2006.

[106] K. Brandenburg and G. Stoll, "Iso/mpeg-1 audio: A generic standard for coding of high-quality digital audio," *Journal of the Audio Engineering Society*, vol. 42, no. 10, pp. 780–792, 1994.

[107] I. E. Commission *et al.*, *Information Technology: Coding of Moving Pictures and Associated Audio for Digital Storage Media at Up to about 1, 5 Mbit/s*. ISO/IEC, 1993.

[108] M. BCl, "An introduction to the psychology of hearing," 1989.

[109] E. Zwicker and H. Fastl, *Psychoacoustics: Facts and models*, vol. 22. Springer Science & Business Media, 2013.

[110] Y. You, *Audio Coding: Theory and Applications*. Springer Science & Business Media, 2010.

[111] R. P. Hellman, "Asymmetry of masking between noise and tone," *Perception & Psychophysics*, vol. 11, no. 3, pp. 241–246, 1972.

[112] M. R. Schroeder, B. S. Atal, and J. Hall, "Optimizing digital speech coders by exploiting masking properties of the human ear," *The Journal of the Acoustical Society of America*, vol. 66, no. 6, pp. 1647–1652, 1979.

[113] P. Noll, "Wideband speech and audio coding," *Communications Magazine, IEEE*, vol. 31, no. 11, pp. 34–44, 1993.

[114] The bark scale: `http://home.ieis.tue.nl/dhermes/lectures/soundperception/04Auditoryl` [Accessed 22 May 2016].

[115] S.-s. Kuo, J. D. Johnston, W. Turin, and S. R. Quackenbush, "Covert audio watermarking using perceptually tuned signal independent multiband phase modulation," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2002.

[116] P. Bassia, I. Pitas, and N. Nikolaidis, "Robust audio watermarking in the time domain," *IEEE Transactions on multimedia*, vol. 3, no. 2, pp. 232–241, 2001.

[117] N. M. Ngo and M. Unoki, "Watermarking for digital audio based on adaptive phase modulation," in *International Workshop on Digital Watermarking*, pp. 105–119, Springer, 2014.

[118] R. Nishimura and Y. Suzuki, "Audio watermark based on periodical phase shift," *J. Acoust. Soc. Jpn*, vol. 60, no. 5, pp. 269–272, 2004.

[119] M. Unoki and D. Hamada, "Audio watermarking method based on the cochlear delay characteristics," in *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP'08 International Conference on*, pp. 616–619, IEEE, 2008.

[120] M. Unoki and R. Miyauchi, "Study on non-blind detection method for digital-audio watermarking technique based on the cochlear delay characteristics," in *FIT2011*, pp. 89–96, 2011.

[121] W. Li, X. Xue, and P. Lu, "Localized audio watermarking technique robust against time-scale modification," *Multimedia, IEEE Transactions on*, vol. 8, no. 1, pp. 60–69, 2006.

[122] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Efficiently self-synchronized audio watermarking for assured audio data transmission," *Broadcasting, IEEE Transactions on*, vol. 51, no. 1, pp. 69–76, 2005.

[123] S. Sun and S. Kwong, "A self-synchronization blind audio watermarking algorithm," in *Intelligent Signal Processing and Communication Systems, 2005. ISPACS 2005. Proceedings of 2005 International Symposium on*, pp. 133–136, IEEE, 2005.

[124] S. Wu, J. Huang, D. Huang, and Y. Q. Shi, "Self-synchronized audio watermark in dwt domain," in *Circuits and Systems, 2004. ISCAS'04. Proceedings of the 2004 International Symposium on*, vol. 5, pp. V–712, IEEE, 2004.

[125] X.-Y. Wang and H. Zhao, "A novel synchronization invariant audio watermarking scheme based on dwt and dct," *Signal Processing, IEEE Transactions on*, vol. 54, no. 12, pp. 4835–4840, 2006.

[126] K. Hiratsuka, K. Kondo, and K. Nakagawa, "On the accuracy of estimated synchronization positions for audio digital watermarks using the modified patchwork algo-

rithm on analog channels," in *Intelligent Information Hiding and Multimedia Signal Processing, 2008. IIHMSP'08 International Conference on*, pp. 628–631, IEEE, 2008.

[127] J. Karnjana, P. H. B. Nhien, S. Wang, N. M. Ngo, and M. Unoki, "Comparative study on robustness of synchronization information embedded into an audio watermarked frame," *IEICE Technical Report*, vol. 115, no. 479, pp. 41–44, 2016.

[128] S. W. Golomb *et al.*, *Shift register sequences*. Aegean Park Press, 1982.

[129] S. A. Gelfand, *Hearing: An introduction to psychological and physiological acoustics*. CRC Press, 2009.

[130] I. Daubechies *et al.*, *Ten lectures on wavelets*, vol. 61. SIAM, 1992.

[131] H.-G. Stark, *Wavelets and signal processing: an application-based introduction*. Springer Science & Business Media, 2005.

[132] M. Sifuzzaman, M. Islam, and M. Ali, "Application of wavelet transform and its advantages compared to fourier transform," 2009.

[133] RWC database: `https://staff.aist.go.jp/m.goto/RWC-MDB/` [Accessed 25 April 2016].

[134] P. Kabal, "An examination and interpretation of itu-r bs. 1387: Perceptual evaluation of audio quality," *TSP Lab Technical Report, Dept. Electrical & Computer Engineering, McGill University*, pp. 1–89, 2002.

[135] J. Karnjana, M. Unoki, P. Aimmanee, and C. Wutiwiwatchai, "An audio watermarking scheme based on singular-spectrum analysis," in *Digital-Forensics and Watermarking*, pp. 145–159, Springer, 2014.

[136] Y.-T. HUANG and H. T. LAWLESS, "Sensitivity of the abx discrimination test," *Journal of sensory studies*, vol. 13, no. 2, pp. 229–239, 1998.

[137] J. Boley and M. Lester, "Statistical analysis of abx results using signal detection theory," in *Audio Engineering Society Convention 127*, Audio Engineering Society, 2009.

[138] G. Casella and R. L. Berger, *Statistical inference*, vol. 2. Duxbury Pacific Grove, CA, 2002.

[139] M. Holmes, A. Gray, and C. Isbell, "Fast svd for large-scale matrices," in *Workshop on Efficient Machine Learning at NIPS*, vol. 58, pp. 249–252, 2007.

[140] P. Drineas, E. Drinea, and P. S. Huggins, "An experimental evaluation of a monte-carlo algorithm for singular value decomposition," in *Advances in Informatics*, pp. 279–296, Springer, 2001.

[141] G. H. Golub and C. F. Van Loan, *Matrix computations*, vol. 3. JHU Press, 2012.

[142] J. N. Kutz, *Data-driven modeling & scientific computation: methods for complex systems & big data*. OUP Oxford, 2013.

[143] Daubechies 1: `http://http://http://wavelets.pybytes.com/wavelet/db1/` [Accessed 17 May 2016].

[144] W. Al-Nuaimy, M. A. El-Bendary, A. Shafik, F. Shawki, A. E. Abou-El-azm, N. A. El-Fishawy, S. M. Elhalafawy, S. M. Diab, B. M. Sallam, F. E. A. El-Samie, *et al.*, "An svd audio watermarking approach using chaotic encrypted images," *Digital Signal Processing*, vol. 21, no. 6, pp. 764–779, 2011.

[145] M. Steinebach, S. Zmudzinski, and T. Bolke, "Audio watermarking and partial encryption," in *Electronic Imaging 2005*, pp. 779–788, International Society for Optics and Photonics, 2005.

[146] A. Lemma, S. Katzenbeisser, M. Celik, and M. Van Der Veen, "Secure watermark embedding through partial encryption," in *Digital Watermarking*, pp. 433–445, Springer, 2006.

[147] Q. Liu, Y. Li, L. Hao, and H. Peng, "Two efficient variants of the rsa cryptosystem," in *2010 International Conference On Computer Design and Applications*, 2010.

[148] J. Jones, "The rsa algorithm," *ACM Communications in Computer Algebra*, vol. 42, no. 1-2, pp. 74–74, 2008.

[149] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems," *Communications of the ACM*, vol. 21, no. 2, pp. 120–126, 1978.

[150] M. Cozzens and S. J. Miller, *The mathematics of encryption: an elementary introduction*, vol. 29. American Mathematical Soc., 2013.

[151] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *Information Theory, IEEE Transactions on*, vol. 47, no. 4, pp. 1423–1443, 2001.

[152] B. Chen and G. W. Wornell, "Quantization index modulation methods for digital watermarking and information embedding of multimedia," *Journal of VLSI signal processing systems for signal, image and video technology*, vol. 27, no. 1-2, pp. 7–33, 2001.

[153] B. Chen and G. W. Wornell, "Preprocessed and postprocessed quantization index modulation methods for digital watermarking," in *Electronic Imaging*, pp. 48–59, International Society for Optics and Photonics, 2000.

[154] IHC: `http://http://www.ieice.org/iss/emm/ihc/` [Accessed 25 April 2016].

[155] Q. X.-m. Z. Hong-bin, "Fragile audio watermarking algorithm for telltale tamper proofing and authentication," *Journal of Electronics and Information Technology*, vol. 8, p. 003, 2005.

[156] P. Yin and H. H. Yu, "A semi-fragile watermarking system for mpeg video authentication," in *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, vol. 4, pp. IV–3461, IEEE, 2002.

[157] L. Boney, A. H. Tewfik, and K. N. Hamdy, "Digital watermarks for audio signals," in *Multimedia Computing and Systems, 1996., Proceedings of the Third IEEE International Conference on*, pp. 473–480, IEEE, 1996.

[158] D. Gruhl, A. Lu, and W. Bender, "Echo hiding," in *International Workshop on Information Hiding*, pp. 295–315, Springer, 1996.

Publications

Journal

[1] Jessada Karnjana, Masashi Unoki, Pakinee Aimmanee, and Chai Wutiwiwatchai, "Singular-Spectrum Analysis of Digital Audio Watermarking with Automatic Parameterization and Parameter Estimation," IEICE Transaction on Information and System, Vol. E99-D, No. 8, Aug., 2016.

[2] Jessada Karnjana, Masashi Unoki, Pakinee Aimmanee, and Chai Wutiwiwatchai, "Audio Watermarking Scheme Based on Singular-Spectrum Analysis and Psychoacoustic Model with Self Synchronization," Recent Advances in Multimedia Watermarking and Forensics, 2016. (Submitted the revision for the 2nd round review)

Lecture Note

[3] Jessada Karnjana, Masashi Unoki, Pakinee Aimmanee, and Chai Wutiwiwatchai, "An Audio Watermarking Scheme based on Singular-Spectrum Analysis," in Digital-Forensics and Watermarking, LNCS 9023, pp. 145–159, 2015.

International Conference

[4] Jessada Karnjana, Pakinee Aimmanee, Masashi Unoki, and Chai Wutiwiwatchai, "An audio watermarking scheme based on automatic parameterized singular-spectrum analysis using differential evolution," in Proc. Asia-Pacific Signal and Information

Processing Association Annual Summit and Conference (APSIPA), pp. 543–551, Hong Kong, Dec., 2015.

[5] Masashi Unoki, Jessada Karnjana, Shengbei Wang, Nhut Minh Ngo, and Ryota Miyauchi, "Comparative Evaluations of Inaudible and Robust Watermarking for Digital Audio Signals," in Proc. The 21st International Congress on Sound and Vibration, China, Jul., 2014.

[6] Jessada Karnjana, Masashi Unoki, Pakinee Aimmanee, and Chai Wutiwiwatchai, "SSA-based Audio-Information-Hiding Scheme with Psychoacoustic Model," Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2016. (Accepted)

[7] Jessada Karnjana, Masashi Unoki, Pakinee Aimmanee, and Chai Wutiwiwatchai, "Tampering Detection in Speech Signals by Semi-Fragile Watermarking Based on Singular-Spectrum Analysis," The 12th International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIHMSP), 2016. (Accepted)

Domestic Conference

[8] Jessada Karnjana, Masashi Unoki,Shengbei Wang, Nhut Minh Ngo, and Ryota Miyauchi, "Comparative Evaluations of Inaudible and Robust Watermarking for Digital Audio Signals," IEICE Tech. Rep. EMM2013-114, pp. 63-68, 2014.

[9] Jessada Karnjana, Masashi Unoki, Pakinee Aimmanee, and Chai Wutiwiwatchai, "Study on Audio Watermarking Scheme Based on Singular-Spectrum Analysis," IEICE Tech. Rep. EMM2014-71, pp. 27-32, 2015.

[10] Jessada Karnjana, Masashi Unoki, "Methods for Concavity Detection in Singular Spectrum," IEICE Tech. Rep., 115(479), pp. 105-110, 2016.

[11] Jessada Karnjana, Pham Hoang Bao Nhien, Shengbei Wang, Nhut Minh Ngo, and Masashi Unoki, "Comparative Study on Robustness of Synchronization Information Embedded into an Audio Watermarked Frame," IEICE Tech. Rep., 115(479), pp. 41-44, 2016.