This is to certify that the
dissertation entitled

SPEECH WATERMARKING THROUGH PARAMETRIC MODELING

presented by

APARNA GURIJALA

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Electrical Engineering

John R. Hellen Jr.
Major Professor's Signature

01 Nov. 2007
Date

SPEECH WATERMARKING THROUGH PARAMETRIC MODELING

By

Aparna Gurijala

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Electrical Engineering

2007

ABSTRACT

SPEECH WATERMARKING THROUGH PARAMETRIC MODELING

By

Aparna Gurijala

Parameter-embedded watermarking of speech is effected through slight perturbations of parametric models of deeply-integrated dynamics of the signal. This research focusses on speech watermarking techniques based on linear-in-parameters speech models. Information is embedded by modifying the linear predictor coefficients of the original speech, subject to fidelity constraints. The modified parameters are used to reconstruct the watermarked speech. Experiments with real speech data are used to assess robustness and other performance properties. A particular example of watermark detector design is discussed and performance tested.

In set-membership filtering (SMF) based parametric watermarking, linear predictor (LP) coefficients of the original speech are modified subject to an objective fidelity constraint. SMF is used to obtain a hyperellipsoidal set of allowable parameter perturbations (i.e., watermarks) subject to a constraint on the error between the watermarked

and original material. This research discusses the robustness of SMF based watermarking to filtering, quantization and combination attacks. An important consideration in watermark robustness is the energy of the watermark signal (difference between watermarked and original signals). Watermarks of higher energy are obtained from perturbed LP coefficients at the boundary of the hyperellipsoidal set. A constrained optimization problem is solved to obtain the best watermarks for filtering and quantization attacks.

Finally, a generalized framework for parametric speech watermarking is presented. In addition to the LP model, other parametric representations such as log area ratio, inverse sine, line spectrum pair, and reflection coefficients are used for speech watermarking. An application of perturbed parameter theory for autoregressive models is presented. The perturbed parameter theory is used to obtain bounds on the perturbation of the stegosignal caused by watermarking.

ACKNOWLEDGMENTS

I would like to thank my faculty advisor Dr. Jack Deller for his guidance, patience, understanding and support throughout my graduate studies at Michigan State University. I sincerely thank him for providing me with an opportunity to conduct research in speech watermarking and for creating a conducive learning environment. I am very grateful to my PhD committee, Drs. Aviyente, Jain, Radha, and Seadle, for their concern, patience, and encouragement. The classroom experience and research discussions with my professors have invaluably contributed to my knowledge and understanding.

I would especially like to thank my family for their love, patience and understanding. My parents always put great on emphasis on education and were willing to support me in every possible way. I would also like to acknowledge my brother, Ashok for his encouragement and his great interest in technology. I would like to thank Ali for his encouragement and confidence in me and for the numerous research discussions we had. I would like to thank Mujahid, Dale, and Margaret for their kindness, encouragement and support throughout my PhD program. Ali, Mujahid, Dale, and Margaret made my stay at MSU a very memorable one.

This work is supported by the National Science Foundation of the

United States under Cooperative Agreement No. IIS-9817485. I would like to acknowledge NSF for their generous support to the National Gallery of Spoken Word project. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Digital media and global access to high-speed computer networks are creating complex copyright issues for owners of legally-protected materials [1]. A response to the unprecedented need to protect intellectual property has been the emergence of an active research effort into digital watermarking technologies. Digital watermarking is the process of embedding data (the *watermark*) imperceptibly into a host signal (the *coversignal*) to create a *stegosignal.* The term "coversignal" is commonly used in watermarking literature [2] to denote the host signal and the term "stegosignal" is borrowed from steganography [3] to represent the watermarked signal. The watermark is typically a pseudo-noise sequence, or a sequence of symbols mapped from a message. A watermark offers copyright protection by providing identifying information which is accessible only to the owner of the material. Only a watermarked version of copyrighted material is released to the public.

When copyright questions arise, the watermark is recovered from the stegosignal as evidence of title. Watermarking has been argued to be an advantageous solution to this modern copyright problem, and there is strong evidence that the practice will be accepted by the courts as proof of title [1].

The design of a watermarking strategy for speech involves the balancing of two principal criteria. First, embedded watermarks must be imperceptible to the listener. That is, the stegosignal must be of high fidelity. Second, watermarks must be robust. That is, they must be able to survive *attacks* [4] - those deliberately designed to destroy or remove them, as well as distortions inadvertently imposed upon the watermarks by technical processes (e.g., compression) or by systemic processes (e.g., channel noise). These *fidelity* and *robustness* criteria are generally competing, as greater robustness requires more watermark energy and more manipulation of the coversignal, which, in turn, lead to noticeable distortion of the original content. Related measures of a watermark's efficacy include *data payload*, the number of watermark bits per unit of time [2]. Another important requirement of a watermarking strategy is its *security*, the inherent protection against unauthorized removal, embedding or detection. A watermarking scheme generally derives its security from secret codes or patterns (*keys*) that are used to embed the watermark. Only a breach of keying strategies should

compromise the security of a watermarking technique; public knowledge of the technical method should not lesson its effectiveness.

The speech watermarking methods described in this dissertation involve *private decoding*, meaning that the coversignal is required for watermark recovery. Private decoding techniques require additional information during watermark detection and recovery. However, among other benefits, this additional information can be used to undo certain attacks and distortion. In private decoding techniques, knowledge of the coversignal at the detector, serves as a registration pattern to undo any temporal or geometric distortions of the stegosignal [2]. For example, in the case of a "cropping attack," wherein speech samples are randomly deleted, a dynamic programming algorithm can be used in conjunction with the coversignal to recover the watermark from the desynchronized stegosignal [5]. Although watermarking schemes involving *public decoding* (coversignal not required for watermark recovery) are applicable in a larger set of applications, techniques involving private decoding can be used for content tracking, broadcast monitoring, and owner identification, in addition to copyright protection.

Robustness requirements of watermarking algorithms are application dependent. Watermarking algorithms are broadly categorized into robust and fragile watermarking algorithms based on the robustness requirements. For a given application, robust watermarking algorithms

are required to survive all intentional attacks and also distortion introduced by normal processing. Fragile watermarking algorithms are required to be selectively robust. For example, in a speech authentication application of watermarking, the embedded fragile watermarks are required to be robust to compression, channel noise, and resampling and fragile to content tampering due to re-embedding and changes to acoustic information. The algorithms presented in this thesis fall under the robust watermarking category and were developed for applications such as content management, broadcast monitoring, and copyright protection.

Watermark embedding techniques vary widely in method and purpose. Watermarks may be additive, multiplicative, or quantization-based, and may be embedded in the time domain, or in a transform domain. Each technical variation tends to be more robust to some forms of attack than to others, and for this and other application-specific reasons, particular strategies may be better-suited to certain tasks. The methods reported in this dissertation are motivated by the particular properties of speech signal [6].

Parametric watermarking is based on manipulation of inear-in-parameters speech models. The linear prediction (LP) model is a special case of linear-in-parameters speech models that can be used for watermarking [6]. Generally speaking, the watermark information is

concentrated in the few LP coefficients during the watermark embedding and recovery processes, while it is dispersed temporally and spectrally otherwise [7]. The watermark recovery process involves least square error (LSE) estimation [8] of modified LP coefficients, and this further contributes to watermark robustness. Parametric watermarking provides sufficient flexibility in terms of watermark selection for a wide range of data payload, robustness, and stegosignal fidelity requirements.

In *set-membership filtering* (SMF) *based parametric watermarking*, LP coefficients of the original speech are modified subject to an objective fidelity constraint. SMF is used to obtain sets of allowable parameter perturbations (i.e., watermarks) subject to a constraint on the error between the watermarked and original material. The robustness of SMF based watermarking to filtering, quantization and combination attacks is studied. An important consideration in watermark robustness is the energy of the watermark signal (difference between watermarked and original signals). The most robust watermark is obtained from perturbed LP coefficients at the boundary of the membership set.[1] A constrained optimization problem is solved to obtain the best watermarks for filtering and quantization attacks.

The application that motivated the present work is the creation of

[1]This phenomenon is discussed below.

the *National Gallery of the Spoken Word* (NGSW), an NSF-sponsored Digital Libraries Initiative II project. The goal of the NGSW effort is the development and management of an extensive on-line repository of spoken word collections, based largely on the renowned Vincent Voice Library. Further information is available at www.lib.msu.edu/vincent/ and in [9].

Owners of copyrighted material are often reluctant to grant permission to post such material on the internet without sufficient assurances that their rights will be protected. Accordingly, a prime interest in the development of the watermarking scheme is the need for robustness to the broadest possible array of attacks. On the other hand, preserving the audio history and authenticity of the NGSW materials requires that robustness not come at the expense of perceptible distortion.

Although the NGSW application places few constraints on computational load, parametric watermarking can be implemented in real-time. Further, since the NGSW is a permanent, large-scale, repository of speech data with a rich meta-data support structure, the association of relatively detailed watermarking information with records in the database is not impractical.

Chapter 2

Background

In the last decade many algorithms have been proposed for multimedia watermarking. Early work emphasized watermarking algorithms that could be universally applied to a wide spectrum of multimedia content, including images, video, and audio. This versatility was deemed conducive to the implementation of multimedia watermarking on common hardware [10]. However, many watermarking applications, including copyright protection for digital speech libraries [11], embedding patient information in medical records [12, 13], or television broadcast monitoring [14], involve embedding information into a single medium. Also, the attacks and inherent processing distortions vary depending on the nature of the data. For example, an attack on watermarked images may involve rotation and translation operations to disable watermark detection. However, such an attack is not applicable to audio data. Watermarking algorithms that are specifically designed for par-

ticular multimedia content can exploit well-understood properties of that content to better satisfy the robustness, fidelity and data-payload constraints. For example, unlike general audio, speech is characterized by intermittent periods of voiced (periodic) and unvoiced (noise-like) sounds. Speech signals are characterized by a relatively narrow bandwidth, with most information below 4 kHz. Also, well-established analytical models for speech production exist [6] which can be exploited in the watermarking process.

2.1 Speech watermarking

Most existing watermarking algorithms for speech can be categorized into either spread-spectrum (SS) or speech synthesis based approaches. SS watermarking [10] is one of the earliest and best-known watermarking algorithms applied to multimedia data. In SS watermarking, a narrowband watermark is embedded into a wideband "channel" that is the coversignal. In the second main approach, watermarks are integrated through speech synthesis. An advantage of integrating watermarking with the coding process [15] is a reduction in computational complexity.

In this work, we adopt a new approach that has both spectrum-spreading and integration-by-synthesis aspects, but which is fundamentally different from the existing approaches. For speech signals, a para-

metric approach is naturally motivated by the extraordinary successes in applying parametric models - in particular, the LP model - in several key speech technology areas. The robustness of the LP model to practical anomalies occurring in coding, recognition, and other applications, suggests that some representation of these parameters might provide an effective basis for embedding durable watermarking data. Parametric watermarking provides sufficient flexibility in terms of watermark selection for a wide range of data payload, robustness, and stegosignal fidelity requirements. In the strategy described here, LP parameters of speech are directly or indirectly modified by an added watermark vector. The stegosignal is constructed by passing the original speech through the modified inverse LP filter and resultant is then added to the prediction residual of the unaltered LP model.

2.1.1 Spread spectrum watermarking

An important contribution of the work of Cox *et al.* [10] is the demonstration that a watermark must be embedded in perceptually significant components of a signal for sufficient robustness to attack. In [10], the DCT is applied to the coversignal and the watermark is embedded in the n (typically 1000) highest magnitude coefficients of the DCT, not including the zero frequency component. Each value of the watermark is drawn independently from a unit normal distribution.

SS watermarking is robust to a wide range of attacks, so it is used as a standard against which to evaluate the robustness of parametric watermarking in this work. For the SS algorithm used to compare performance in this research, the stegosignal $\{\breve{y}_j\}_{j=1}^{N}$ is obtained by adding the watermark sequence $\{g_i\}_{i=1}^{1000}$ to the 1000 largest DCT coefficients of the coversignal of 1 s duration.

$$\breve{Y}_i = Y_i + Y_i \lambda g_i, \tag{2.1}$$

where each g_i is independently drawn from $\mathcal{N}(0,1)$, and Y_i and $\breve{Y}_i$ are the i^{th} largest DCT coefficients of the cover and stegosignals, respectively. The λ parameter controls the stegosignal fidelity and is adjusted to satisfy a desired fidelity constraint.

In SS signaling [16, 17], the watermark message is first modulated by a lowpass filtered pseudo-noise sequence. The resulting sequence is shaped by the LP spectrum of the coversignal, before being added to the coversignal. The latter measure reduces perceptual distortion. The watermark receiver whitens the stegosignal using the inverse LP filter. The watermark receiver requires perfect synchronization between the whitened stegosignal and the pseudo-noise spreading sequence. These techniques have been tested in low noise environments such as in the presence of additive white Gaussian noise with a 20 dB SNR. However, it is not known how such algorithms will perform under more challeng-

ing channel conditions, or when subjected to deliberate attacks like cropping, filtering, or the addition of colored noise.

2.1.2 Watermarking integrated with speech synthesis

In the approach by Hatada *et al.* [18], line spectrum pairs (LSP) [6] are extracted from short-term segments of the coversignal. The LSP parameters are selected because they correlate well with the formant location [18]. Codebook vectors are created by applying a clustering algorithm to the extracted LSPs. Watermarked codebook vectors are obtained by modifying the frequency components of the original codebook vectors. The LSPs of a particular frame are quantized by either the watermarked or original codebooks depending on whether the frame is to be watermarked or not. The stegosignal is synthesized using the watermarked LSPs.

Even in the absence of watermarking, the LSPs of the original speech and those of the synthesized speech are different. In the presence of watermarking, the difference between the original and extracted LSPs will be even more substantial. Thus watermark detection is affected even in the absence of an attack. Hence, to preserve the watermark information as accurately as possible, it is necessary that the speech frames used for embedding watermark data have very small LSP differences with respect to the synthesized speech.

2.1.3 Pitch and duration modification for watermarking

Celik *et al.* [19] propose a speech watermarking algorithm for semi-fragile authentication applications. In the case of semi-fragile watermarking, robustness to selective manipulations or attacks is desired. Celik *et al.* use pitch and duration modification of quasi-periodic speech phonemes as the features for semi-fragile watermarking. The significance of these features makes them suitable for watermarking and the variability of these features facilitates imperceptible data embedding. A quantization index modulation scheme is used to embed watermark bits into these features.

The coversignal is segmented into phonemes. A phoneme is a fundamental unit of speech that conveys linguistic meaning [6]. Certain classes of phonemes such as vowels, semivowels, diphthongs, and nasals are quasi-periodic in nature. The periodicity is characterized by the fundamental frequency or the pitch period. The pitch synchronous overlap and add (PSOLA) algorithm is used to parse the coversignal and to modify the pitch and duration of the quasi-periodic phonemes [20]. The pitch periods ($\acute{\rho}_p$) are determined for each segment of the parsed coversignal. The average pitch period is then computed for each segment,

$$\acute{\rho}_{avg} = \sum_{p=1}^{P} \frac{\acute{\rho}_p}{P}$$

The average pitch period is modified to embed the m^{th} watermark bit ($\acute{\omega}_m$) by using dithered quantization index modulation [21],

$$\acute{\rho}_{avg}^{\acute{\omega}} = Q_{\acute{\omega}}(\acute{\rho}_{avg} + \acute{\eta}) - \acute{\eta}$$

where $Q_{\acute{\omega}}$ is the selected quantizer and $\acute{\eta}$ is the pseudo-random dither value. Pitch periods is then modified such that,

$$\acute{\rho}_{p}^{\acute{\omega}} = \acute{\rho}_{p} + (\acute{\rho}_{avg}^{\acute{\omega}} - \acute{\rho}_{avg})$$

The PSOLA algorithm is used to concatenate the segments and synthesize the stegosignal. The duration of the segments is modified for better reproduction of the stegosignal. As necessitated by authentication applications, watermark detection does not require the original speech. At the detector, the procedure is repeated and the modified average pitch values are determined for each segment. Using the modified average pitch values, the watermark bits are recovered. The algorithm is robust to distortions caused by low-bit-rate speech coding. This is because it uses features that are preserved by low-bit-rate speech coders such as QCELP, AMR, and GSM-06.10 [22]. Robustness to coding and compression is necessary for authentication applications. On the other hand, the fragile watermarking algorithm is designed to detect malicious operations such as re-embedding and changes to acoustic in-

formation (e.g., phonemes).

2.2 Set-membership filtering

The set-membership filtering (SMF) concept was first published by Gollamudi *et al.* [23], and was more recently proposed as an innovative solution to the design of channel equalizers for digital communication by Nagaraj *et al.* [24]. SMF can be viewed as a reformulation of the broadly-researched class of algorithms concerned with *set-membership identification* (e.g. [25, 26]). The application of SMF to parametric speech watermarking in demonstrated in Chapter 4.

2.2.1 Overview of SMF

The SMF problem is stated as follows:

SMF PROBLEM. *Given a sequence* $\left\{\mathbf{x}_\tau \in \mathbb{R}^M\right\}_{\tau=1}^{t}$ *of observations, a "desired" sequence* $\left\{z_\tau \in \mathbb{R}\right\}_{\tau=1}^{t}$, *and a sequence of error "tolerances"* $\left\{\gamma_\tau\right\}_{\tau=1}^{t}$ (*frequently constant with* τ), *find the the exact feasibility set at time* t, $\mathcal{P}_t \subseteq \mathbb{R}^M$ *which includes all vectors (filters),* $\theta \in \mathbb{R}^M$, *satisfying*

$$\mathcal{P}_t = \left\{\theta \,\middle|\, |z_\tau - \theta^T \mathbf{x}_\tau| < \gamma_\tau \quad \text{for} \quad \tau \in [1, t]\right\}. \tag{2.2}$$

Note that when γ_t is constant with t, say $\gamma_t = \gamma$, then we may write

$$\mathcal{P}_t = \{\theta \,|\, \|\mathbf{z} - \bar{\mathbf{z}}\|_\infty < \gamma\}\,. \tag{2.3}$$

in which $\mathbf{z}$ is the t-vector with i^{th} element z_i, and $\bar{\mathbf{z}}$ is the t-vector with i^{th} element $\mathbf{x}_i^T\theta$.

The SMF problem is solved using a series of recursions which return at iteration t an hyperellipsoidal membership set, say $\mathcal{E}_t \supset \mathcal{P}_t$, and the ellipsoid's center, say θ_t. The recursions execute an optimization strategy designed to tightly bound $\mathcal{P}_t$ by $\mathcal{E}_t$ in some sense. Accordingly, the broad class of algorithms employed in the SMF problem are often called the *optimal bounding ellipsoid (OBE)* algorithms [25, 26]. The OBE algorithm used in the SMF based parametric watermarking algorithm is called the *set-membership—weighted recursive least squares* (SM-WRLS) algorithm, but the choice of OBE methods is somewhat arbitrary for the present application.

2.2.2 Set-membership weighted recursive least squares

This section presents an overview of the SM-WRLS algorithm for filtering and identification applications. The SM-WRLS algorithm is used in the SMF based parametric watermarking algorithm. In the SMF framework it is assumed that there is an observation sequence $\{\mathbf{x}_t\}_{t=1}^{\infty}$, a "desired" sequence $\{z_t\}_{t=1}^{\infty}$, and a sequence of error tolerances

$\{\gamma_t\}_{t=1}^{\infty}$ [25]. The feasibility set at time t, $\mathcal{P}_t$, includes all θ, such that

$$\bar{z}_\tau = \theta^T \mathbf{x}_\tau \text{ subject to } |z_\tau - \bar{z}_\tau| < \gamma_\tau \text{ for } \tau = 1, 2, \ldots, t. \tag{2.4}$$

Let $\bar{\mathbf{X}}_t$ be the $t \times M$ matrix with the i^{th} row $\sqrt{\lambda_i}\mathbf{x}_i^T$ and let $\mathbf{z}_t$ be the t-vector with the i^{th} element $\sqrt{\lambda_i} z_i$, where $\{\sqrt{\lambda_\tau}\}_{\tau=1}^{t}$ are a set of error minimization weights. Then the covariance matrix is given by $\mathbf{C}_x = \bar{\mathbf{X}}_t^T \bar{\mathbf{X}}_t$ and $\mathbf{c}_{xz} = \bar{\mathbf{X}}_t^T \mathbf{z}_t$. The algorithmic steps involved in implementing SM-WRLS for either identification or filtering applications are given in Table 2.1 [27].

2.3 Lagrange Multipliers

The method of Lagrange Multipliers is a common approach for solving constrained optimization problems [28]. The method of Lagrange Multipliers is used to obtain optimal watermarks from the membership set for a given attack on the stegosignal. In a constrained optimization problem, a function needs to be maximized or minimized subject to certain conditions or constraints. A constrained optimization problem with the variable $x \in \mathbb{R}^n$ is characterized by an objective function $f_o(x)$, inequality constraint functions $f_i(x)$ and equality con-

straint functions $h_j(x)$.

$$\begin{aligned} &\min/\max \quad f_o(x) \\ &\text{subject to} \quad f_i(x) \leq 0, \text{ for } i = 1, 2, \ldots, p_1 \\ &\qquad\qquad h_j(x) = 0, \text{ for } j = 1, 2, \ldots, p_2 \end{aligned} \tag{2.5}$$

In the method of Lagrange multipliers, the constraint functions are taken into account by augmenting a weighted combination of the constraint functions to the objective function [28]. That is, the Lagrangian $L(x, \breve{\lambda}, \breve{\nu})$ is,

$$L(x, \breve{\lambda}, \breve{\nu}) = f_o(x) + \sum_{i=1}^{p_1} \breve{\lambda}_i f_i(x) + \sum_{j=1}^{p_2} \breve{\nu}_j h_j(x) \tag{2.6}$$

where $L : \mathbb{R}^n \times \mathbb{R}^{p_1} \times \mathbb{R}^{p_2} \rightarrow \mathbb{R}$, $\{\breve{\lambda}_i\}_{i=1}^{p_1}$ are the Lagrange multipliers associated with the inequality constraints, and $\{\breve{\nu}_j\}_{j=1}^{p_2}$ are the Lagrange multipliers associated with equality constraints. The Lagrange multipliers are nonzero and those associated with inequality constraints are also nonnegative. The method of Lagrange multipliers converts the constrained optimization problem into an unconstrained one with $n + p_1 + p_2$ variables.

The maxima or minima of the constrained optimization problem occur when the gradient of the Lagrangian is zero, $\nabla L(x, \breve{\lambda}, \breve{\nu}) = 0$.

That is,

$$\nabla_x L(x, \breve{\lambda}, \breve{\nu}) = 0 \iff \nabla_x f_o = -\left(\sum_{i=1}^{p_1} \breve{\lambda}_i \nabla_x f_i + \sum_{j=1}^{p_2} \breve{\nu}_j \nabla_x h_j\right) \quad (2.7)$$

and

$$\nabla_{\breve{\lambda}_i} L(x, \breve{\lambda}, \breve{\nu}) = 0 \iff f_i = 0, \text{ for } i = 1, 2, \ldots, p_1 \quad (2.8)$$

$$\nabla_{\breve{\nu}_j} L(x, \breve{\lambda}, \breve{\nu}) = 0 \iff h_j = 0, \text{ for } j = 1, 2, \ldots, p_2. \quad (2.9)$$

It can also be observed that,

$$\frac{\partial L}{\partial f_i} = \breve{\lambda}_i \text{ and } \frac{\partial L}{\partial h_j} = \breve{\nu}_j. \quad (2.10)$$

Equations (2.7)-(2.10) are used to obtain the optimal value of x. In order to use the method of Lagrange multipliers, the objective and constrained functions are not required to be convex.

Table 2.1: SM-WRLS algorithm

Initialization:

$\mathbf{C}_0^{-1} = \mathbf{P}_0 = \epsilon^{-1}\mathbf{I}$, where ϵ is small
$\theta_0 = 0$
$\lambda_1 = 1$
$\kappa_0 = [\theta_1^T \mathbf{x}_1]^2 + z_1^2$, computed after step 3 for $\tau = 1$

Recursion: For $\tau = 1, 2, \ldots, t$

1 $G(\tau)$ and $\epsilon_{\tau-1}(\tau)$ are updated.

$$G(\tau) = \mathbf{x}_\tau^T \mathbf{P}_{\tau-1} \mathbf{x}_\tau \text{ where } \mathbf{P}_\tau = \mathbf{C}_\tau^{-1}$$

$$\epsilon_{\tau-1}(\tau) = z_\tau - \theta_{\tau-1}^T \mathbf{x}_\tau$$

2 Skip step 2 if $\tau = 1$. The original λ_τ^* is computed by finding a positive root of the following quadratic equation.

$$F(\lambda) = 0 = \begin{cases} \{(M-1)G^2(\tau)\}\lambda^2 \\ +\{[2M - 1 + \gamma_n \epsilon_{\tau-1}^2(\tau)] - \kappa_{\tau-1}\gamma_\tau G(\tau)\}G(\tau)\lambda \\ +\{M[1 - \gamma_\tau \epsilon_{\tau-1}^2(\tau)] - \kappa_{\tau-1} G(\tau)\gamma_\tau\} \end{cases}$$

If there are two positive roots, then the larger one is used.

3 Skip step 3 if $\tau = 1$. If $\lambda_\tau^* \leq 0$, set $\mathbf{P}_\tau = \mathbf{P}_{\tau-1}$, $\theta_\tau = \theta_{\tau-1}$, $\kappa_\tau = \kappa_{\tau-1}$ then go to step 5. Otherwise continue with step 4.

4 Update $\mathbf{P}_\tau$, θ_τ, and κ_τ.

$$\mathbf{C}_\tau^{-1} = \mathbf{P}_\tau = \mathbf{P}_{\tau-1} - \lambda_\tau \frac{\mathbf{P}_{\tau-1}\mathbf{x}_\tau \mathbf{x}_\tau^T \mathbf{P}_{\tau-1}}{1 + \lambda_\tau G(\tau)}$$

$$\theta_\tau = \theta_{\tau-1} + \lambda_\tau \mathbf{P}_\tau \epsilon_{\tau-1}(\tau)\mathbf{x}_\tau$$

$$\kappa_\tau = \kappa_{\tau-1} + \frac{\lambda_\tau}{\gamma_\tau} - \frac{\lambda_\tau \epsilon_{\tau-1}^2(\tau)}{1 - \lambda_\tau G(\tau)}$$

5 If $\tau < t$, increment τ and return to Step 1.

Chapter 3

Parametric Speech Watermarking in the LP Domain

3.1 Introduction

The general parametric watermarking algorithm is formulated in the following way. Let $\{y_n\}$ denote the coversignal, and let $\{\tilde{y}_n\}$ be the ultimate stegosignal. Each of these is assumed to be a real scalar sequence over discrete-time n. It is assumed that the signals are generated according to operations of the form [29]

$$y_n = \phi_\pi(\xi_n, x_n, n) \quad \text{and} \quad \tilde{y}_n = \Phi_{\tilde{\pi}}(\tilde{\xi}_n, \tilde{x}_n, n), \tag{3.1}$$

in which $\{\xi_n\}$, $\{\tilde{\xi}_n\}$, $\{x_n\}$, and $\{\tilde{x}_n\}$ are measurable vector-valued random sequences. The operator ϕ is parameterized by a set π, the alteration of which (to create parameter set $\tilde{\pi}$) is responsible for changing

the operator ϕ to Φ and the sequences $\{\xi_n\}$ and $\{x_n\}$ into their "tilded" counterparts.

3.1.1 An algorithm for LP parametric watermarking

In the present study, the coversignal is assumed to be generated by a LP model,

$$y_n = \sum_{i=1}^{M} a_i y_{n-i} + \xi_n, \tag{3.2}$$

a special case of the first equation in (3.4). The "true" model is determined by standard LP analysis of a (long) frame selected for watermarking [6]. The sequence $\{\xi_n\}$ is the prediction residual associated with the estimated model. The duration of the FIR linear predictor is naturally based on the assumed order of the LP model, M, used to initially parameterize the speech. The stegosignal is constructed using the FIR filter model

$$\tilde{y}_n = \sum_{i=1}^{M} \tilde{a}_i y_{n-i} + \xi_n, \tag{3.3}$$

where $\{\tilde{a}_i\}$ represents a deliberately perturbed version of the "true" set $\{a_i\}$. The algorithmic steps of the LP parameter-embedded watermarking procedure appear in Table 3.1. Numerous ways in which parametric modification can be effected – including indirectly through changes to other speech parameters such as log area ratio (LAR) values or parcor values – are discussed further in Chapter 5.

Table 3.1: Watermark embedding algorithm

Let $\{y_n\}_{n=-\infty}^{\infty}$ denote a coversignal, and let $\{y_n\}_{n=n_k}^{n'_k}$ be the k^{th} of K speech frames to be watermarked. Then: For $k = 1, 2, \ldots, K$

1 Using the "autocorrelation method" (e.g., [6, Ch. 5]), derive a set of LP coefficients of order M, say $\{a_i\}_{i=1}^{M}$, for the given frame.

2 Use the LP parameters in an *inverse filter* configuration to obtain the prediction residual on the frame, $\left\{\xi_n = y_n - \sum_{i=1}^{M} a_i y_{n-i}\right\}_{n=n_k}^{n'_k}$.

3 Modify the LP parameters in some predetermined way to produce a new set, say $\{\tilde{a}_i\}_{i=1}^{M}$. The modifications to the LP parameters (or, equivalently, to the autocorrelation sequence or line spectrum pairs, etc.) comprise the watermark.

4 Use the modified LP parameters as a (suboptimal) predictor of the original sequence, adding the residual obtained in Step 2 above at each n, to resynthesize the speech over the frame, $\left\{\tilde{y}_n = \sum_{i=1}^{M} \tilde{a}_i y_{n-i} + \xi_n\right\}_{n=n_k}^{n'_k}$. (To the extent that the watermark represents only small perturbations to the original LP parameters, the resynthesized result is a pointwise approximation to the coversignal over the same time frame.)

5 The sequence $\{\tilde{y}_n\}_{n_k}^{n'_k}$ is the k^{th} frame of the watermarked speech (stegosignal).

Next k.

When watermark embedding involves direct modification of LP coefficients, the embedding process can be interpreted as a digital filter design problem. Equation (3.3) can be rewritten as

$$\tilde{y}_n = \sum_{i=1}^{M} a_i y_{n-i} + \sum_{i=1}^{M} \omega_i y_{n-i} + \xi_n, \tag{3.4}$$

wherein, the watermark sequence $\{\omega_i\}_{i=1}^{M}$ constitutes the impulse response of an M^{th}-order non-recursive filter. This filtered version of original speech incorporates the watermark information. Non-recursive filters are inherently stable and less sensitive to quantization errors. The watermark signal, $w_n = \sum_{i=1}^{M} \omega_i y_{n-i} = \tilde{y}_n - y_n$ has a spectrum determined by the watermark coefficients and the coversignal. For example, the watermark spectrum can be designed to have predominantly lowpass, highpass or mid-band energy.

It is important to understand a key difference in the way LP modeling is applied in this watermarking application relative to its conventional deployment in speech coding and recognition. In these prevalent applications, the goal is to find a set of LP coefficients that optimally model quasi-stationary regions of speech. In parametric watermarking, the LP model is used as a device to parameterize long intervals of nonstationary speech *without the intention of properly parameterizing stationary dynamics in the waveform.* Rather, the parameters are derived according to the usual optimization criterion – to minimize the

total energy in the residual [6, Ch. 5] – with the understanding that the aggregate time-varying dynamics will be distributed between the long-term parametric code and the residual sequence.

3.1.2 Recovering LP parameter-embedded watermarks

The algorithm for recovering the watermark from the stegosignal appears in Table 3.2. An important step in the recovery process is the least square error (LSE) estimation of the modified watermark coefficients, $\{\tilde{a}_i\}_{i=1}^{M}$, which is executed as follows. Let us consider a length N frame of the coversignal and rewrite the stegosignal generation equation (3.3) as

$$d_n = \sum_{i=1}^{M} \tilde{a}_i y_{n-i} = \tilde{\mathbf{a}}^T \mathbf{y}_n \quad \text{with} \quad d_n = \tilde{y}_n - \xi_n. \tag{3.5}$$

In principle, the system of equations (3.5) taken over N samples, $n = 1, 2, \ldots, N$, is noise free and can be solved for $\tilde{\mathbf{a}}$ without error using any consistent subset of M equations. For generality, to smooth round-off and other errors, and to support further developments, we pose the problem as an attempt to compute the LSE linear estimator of the "output" signal, d_n, given observations $\mathbf{y}_n$. The following normal equations are solved,

$$\mathbf{C}_y \tilde{\mathbf{a}} = \mathbf{c}_{yd} \tag{3.6}$$

Table 3.2: Watermark recovery algorithm

For $k = 1, 2, \ldots, K$

1 Subtract residual frame $\{\xi_n\}_{n_k}^{n'_k}$ from the stegosignal frame $\{\tilde{y}_n\}_{n_k}^{n'_k}$. This results in an estimate of the modified predicted speech, $\{d_n = \tilde{y}_n - \xi_n\}_{n_k}^{n'_k}$.

2 Estimate the *modified* LP coefficients $\{\tilde{a}_i\}_1^M$ by computing the least-square-error solution, say $\{\hat{\tilde{a}}_i\}_1^M$, to the overdetermined system of equations: $d_n \approx \sum_{i=1}^M \alpha_i y_{n-i}$, $n = n_k, \ldots, n'_k$.

3 Use the parameter estimates from Step 2 to derive the corresponding watermark values.

Next k.

in which $\mathbf{C}_y = \sum_{n=1}^N \mathbf{y}_n \mathbf{y}_n^T = \mathbf{Y}_N \mathbf{Y}_N^T$ and $\mathbf{c}_{yd} = \sum_{n=1}^N \mathbf{y}_n d_n = \mathbf{Y}_N \mathbf{d}_1^N$, where

$$\mathbf{Y}_N = \begin{bmatrix} \mathbf{y}_N & \mathbf{y}_{N-1} & \cdots & \mathbf{y}_1 \end{bmatrix} \in \mathbb{R}^{M \times N} \tag{3.7}$$

$$\mathbf{d}_1^N = \begin{bmatrix} d_N & d_{N-1} & \cdots & d_1 \end{bmatrix}^T \in \mathbb{R}^{N \times 1}. \tag{3.8}$$

The LSE method is based on time averages, and its performance depends on the frame length used in the estimation [8]. In the stegosignal, the watermark information is distributed in time and is present as the watermark signal $\{w_n\}_{n=1}^N$. During recovery the watermark information is concentrated in a few coefficients $\{\omega_i\}_{i=1}^M$ derived from an estimate of the modified LP coefficients.

3.1.3 Perceptual aspects of LP parametric watermarking

The watermark embedding process can be interpreted as (i) a modification to the LP model or similarly derived models, plus (ii) FIR filtering. This section deals with the perceptual benefits of parametric watermarking and the constraint used in this research to objectively quantify stegosignal fidelity. Listening tests were also conducted on the watermarked speech file available at the website [30]. The results of these tests are discussed in detail in Section 3.2.3.

Echo embedding interpretation

The LP parametric watermarking agorithm can be interpreted as the addition of M echoes of small amplitudes and scales. The echoes are delayed by M units or less. Typically, echoes of delay 20 mS or less are imperceptible. Also, since the echoes are scaled by much smaller valued watermark coefficients, the louder coversignal masks some components of the echoes. It should be noted that the technique differs from the echo hiding method of Gruhl *et al.* [31], in which binary "one" and "zero" information is encoded in the offset and delay parameters of the echo and not in the echo amplitude.

Stegosignal fidelity

Fidelity is a measure of perceptual similarity between the coversignal and the stegosignal. The watermarking process must not affect the fidelity of the speech beyond an application-dependent standard. A simple and mathematically tractable measure of fidelity is the signal-to-noise ratio (SNR), or, in the present context, *coversignal-to-watermark* ratio (CWR), defined as,

$$\text{CWR} = 10\log_{10}\frac{E_y}{E_w} = 10\log_{10}\frac{\sum_{n=1}^{N} y_n^2}{\sum_{n=1}^{N} w_n^2}, \tag{3.9}$$

where $w_n = [\tilde{y}_n - y_n]$. The CWR averages the relative distortion energy of the coversignal over time and frequency. However, CWR is a poor measure of speech fidelity for a wide range of distortions. The CWR is not related to any subjective attribute of speech fidelity, and it weighs the time domain errors equally [6]. A better measure of speech fidelity can be obtained if the CWR is measured and averaged over short speech frames. The resulting fidelity measure is known as *segmental* CWR [6], defined as,

$$\text{CWR}_{\text{seg}} = \frac{1}{K}\sum_{j=1}^{K} 10\log_{10}\left[\sum_{l=k_j-L+1}^{k_j}\frac{y_l^2}{[\tilde{y}_l - y_l]^2}\right], \tag{3.10}$$

where, $k_1, k_2, ..., k_K$ are the end-times for the K frames, each of which is length L. The segmentation of the CWR assigns equal weight to the

loud and soft portions of speech. For computing CWR_{seg}, the duration of speech frames is typically $15 - 25$ ms with frames of 15 ms used for the experimental results presented in this chapter.

Some of the other objective measures of speech quality include the Itakura distance [6], the weighted-slope spectral distance and the cepstral distance. According to Wang *et al.* [32], CWR_{seg} is a much better correlate to the auditory experience than the other objective measures discussed above. A simple way to control the fidelity of the stegosignal is to scale the watermark vector, ω, by a constant, say λ, before adding it to the original LP parameters (Step 3 of Table 3.1).

3.1.4 Security issues

A watermark's security refers to its ability to withstand attacks designed for unauthorized removal, detection or embedding. A watermarking technique must not rely on the secrecy of the algorithm for its security. In parametric watermarking, a copy of the coversignal is required for watermark recovery. The LP parameters of the stegosignal are different from the modified LP values obtained by adding the watermark vector to the LP parameters of the coversignal. An attacker has access to the stegosignal and not to the coversignal, prediction residual, frame length, and LP model order used for watermarking. Since parametric watermarking involves the alteration of deeply-integrated

characteristics of speech signals, the embedded watermark information is not easily determined from the resulting stegosignals. The security of the present watermarking technique can also be further enhanced by randomly selecting the speech frames to be watermarked, using a different LP model order for each watermarked frame (model order also depends on the fidelity constraint), and by embedding psuedo random watermark patterns. The LP parameters of the stegosignal can be easily obtained.

$$\tilde{y}_n = \sum_{k=1}^{k} \hat{\tilde{a}}_k \tilde{y}_{n-k} + \hat{\tilde{\xi}}_n, \tag{3.11}$$

where K is the LP model order selected by the attacker. However, the LP parameters ($\{\hat{\tilde{a}}_k\}$) of the stegosignal are different from the modified LP coefficients $\{\tilde{a}_i\}$ and also, $\{\hat{\tilde{\xi}}_n\}$ is different from the prediction residual $\{\xi_n\}$) associated with the coversignal, even if $K = M$.

Impact of ambiguity attacks: Ambiguity attacks are of concern to both private and public watermarking techniques [33]. In an ambiguity attack, counterfeit watermarks are identified or created by an attacker in the stegosignal using a different watermarking scheme. The attacker recovers his or her watermark from the stegosignal, claims rightful ownership of the protected signal and succeeds in causing ambiguity about the "true" owner of the stegosignal. According to Craver *et al.*, two necessary conditions for robustness to ambiguity attacks are non-invertibility and non-quasi-invertibility [33]. For a watermarking

technique to be non-invertible, it is essential that the mapping from the watermarked signal $\{\tilde{y}_n\}$ to $\{\acute{\omega}_i\}$ and $\{\acute{y}_n\}$ does not exist; where $\{\acute{\omega}_i\}$ is the watermark carved out by the attacker and $\{\acute{y}_n\}$ is the fake original created by the attacker. Non-quasi-invertibility is a much more stringent requirement. For a watermarking technique to be non-quasi-invertible, it should be impossible for an attacker to create $\{\acute{\omega}_i\}$ and $\{\acute{y}_n\}$ from $\{\hat{\tilde{y}}_n\}$, which is perceptually similar to $\{\tilde{y}_n\}$ and such that $\{\acute{\omega}_i\}$ still exists in $\{y_n\}$, the true original. It is shown below that parametric speech watermarking is non-invertible.

For an algorithm to be invertible, it should be possible for an attacker to create a fake coversignal and a fake watermark from the stegosignal [equation (3.3)].

$$\tilde{y}_n = \sum_{k=1}^{K} \tilde{\acute{a}}_k \acute{y}_{n-k} + \acute{\xi}_n = \sum_{k=1}^{K} (\acute{a}_k + \acute{\omega}_k)\acute{y}_{n-k} + \acute{\xi}_n, \qquad (3.12)$$

where K is the model order selected by the attacker, $\acute{y}_n = \sum_{k=1}^{K} \acute{a}_i \acute{y}_{n-i} + \acute{\xi}_n$ is the fake coversignal, and $\{\acute{\xi}_n\}$ is the corresponding minimum MSE prediction residual.

An attacker can easily compute the LP coefficients and the prediction residual associated with the stegosignal. Obviously, equation (3.11) cannot be substituted by the attacker as the model for the fake coversignal and fake watermark sequence, since the fake coversignal will be

the same as the stegosignal. On the other hand, an attacker can add a sequence $\{\nu_n\}$ to the stegosignal and then compute the LP coefficients and prediction residual.

$$\tilde{y}_n + \nu_n = \sum_{k=1}^{K} \tilde{a}_k(\tilde{y}_{n-k} + \nu_{n-k}) + \acute{\xi}_n = \sum_{k=1}^{K} \tilde{a}_k \acute{y}_{n-k} + \acute{\xi}_n, \qquad (3.13)$$

The attacker can then subtract, $\{\nu_n\}$ from $\{\acute{\xi}_n\}$ to obtain the stegosignal.

$$\tilde{y}_n = \sum_{k=1}^{K} \tilde{a}_k \acute{y}_{n-k} + (\acute{\xi}_n - \nu_n) = \sum_{k=1}^{K} (\acute{a}_k + \acute{\omega}_k)\acute{y}_{n-k} + (\acute{\xi}_n - \nu_n), \quad (3.14)$$

Comparing equations (3.12) and (3.14), the fake coversignal, is given by $\acute{y}_n = \sum_{k=1}^{K} \acute{a}_k \acute{y}_{n-k} + (\acute{\xi}_n - \nu_n)$, where $\acute{\xi}_n - \nu_n$ is the prediction residual. However, from equation (3.13), the fake coversignal is $\acute{y}_n = \tilde{y}_n + \nu_n$ and the minimum MSE prediction residual is $\acute{\xi}_n$, which is different from $(\acute{\xi}_n - \nu_n)$ and hence this is a contradiction. Thus it will be impossible for an attacker to invert the embedding process starting with the stegosignal.

3.1.5 A detection algorithm for LP parametric watermarking

A common approach to watermark detection employs classic binary decision theory. The hypotheses are $H_0 : I_R = I$ and $H_1 : I_R = I + \mathbf{w}$, where I_R is the received signal, I is the original signal and $\mathbf{w}$ is the watermark signal [34,35]. A Bayesian or Neyman-Pearson paradigm

is followed in deriving the detection thresholds. For image watermarking, the image DCT coefficients are modeled as generalized Gaussian in distribution [34, 36]. These approaches do not consider the effect of noise while deriving the detection threshold. Several watermark detectors are based on correlation detection in the time or in the DCT domain [37, 38]. That is, the correlation between the original and recovered watermarks or the correlation between the original watermark and recovered signal is compared against a threshold. Correlation detectors are optimal when the watermark and noise are jointly Gaussian, or, in case of blind detectors, when the watermarked signal and noise are jointly Gaussian. For example, the detector presented in [2, Ch. 6], assumes that the detector output for each bit is Gaussian distributed. This is true for watermark patterns that are spectrally white, but this is not the case with the watermark signal in parametric watermarking. Hence there is a need to design a watermark detector in the parameter domain.

This section describes a watermark detector for LP parametric watermarking [39]. The stegosignal is distorted by additive white or colored Gaussian noise in the time domain. The watermarks are comprised of (eight) non-binary orthogonal vectors of length eight. Each of these eight vectors can be mapped to a unique symbol. For example, each vector can be interpreted as a particular integer from the set

$\{0, 1, 2, 3, 4, 5, 6, 7\}$. The watermark may be composed of many such integers or symbols. In the examples in this paper, each orthogonal watermark vector (symbol), is embedded into 0.125 seconds of speech sampled at 16 kHz, resulting in a bit rate of 24 bits per second (bps). The watermark vector is added to the coefficients of an eighth-order LP model. The length of the watermark vector (and hence the predictor model order) and the duration of speech frame can be selected arbitrarily, subject to constraints on stegosignal fidelity. These constraints include an upper limit on the predictor model order, and a need to use FIR models of small order for short speech frames (~500 samples).

Extensive experimentation by the author has shown that noise in the parameter domain, caused by stegosignal exposure to additive noise, is well-modeled by a Gaussian distribution. Figure 3.1(a) shows a typical noise distribution in the LP domain when white noise (SNR 15 dB) is added to the stegosignal. The noise distribution for Fig. 3.1(a) was obtained by conducting 1000 experiments, involving a stegosignal of 1 s duration watermarked at CWR_{seg} of 7 dB using a watermark message consisting of eight orthogonal vectors, each vector embedded into 0.125 seconds (2000 samples) of speech. When white Gaussian noise was added to the stegosignal, the noise effects on a particular watermark coefficient could be approximated as independent and identically distributed (i.i.d) Gaussian noise. The LP noise associated with

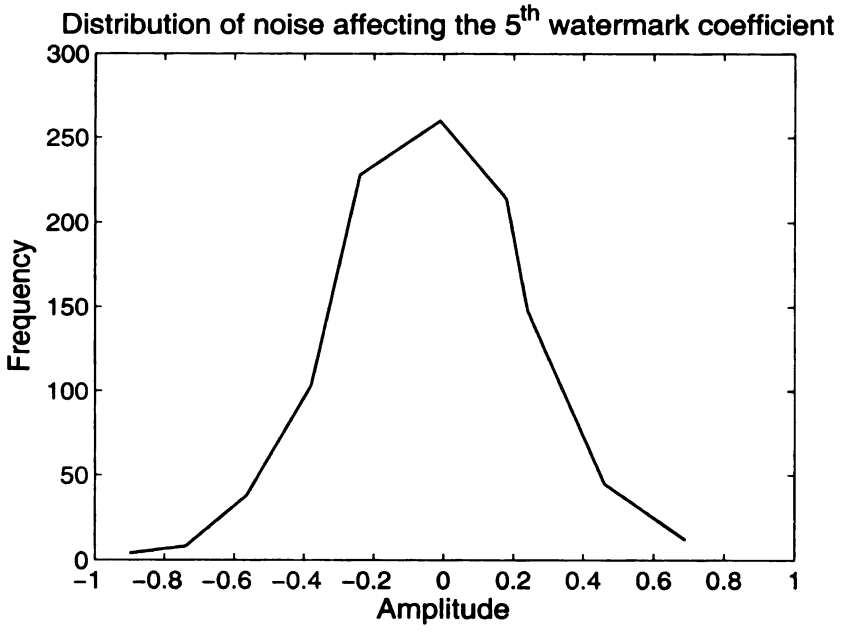
a particular watermark coefficient was uncorrelated with the LP noise for other watermark coefficients. The parameter noise samples were also uncorrelated with the corresponding LP coefficients. It should be noted that when the stegosignal is subjected to additive white noise, the parameter noise asymptotically tends to zero as $N \rightarrow \infty$ (discussed further in Section 3.2.3) and is of very low power. However, the noise generated using the "randn" function in matlab, is not ideal white noise.

The parameter noise distribution of the stegosignal plus colored noise is similar to that shown in Fig. 3.1(b). Colored noise was generated by lowpass filtering white noise using an IIR Butterworth filter. The LP noise affecting any given watermark coefficient was found to be i.i.d. Gaussian. However, a realization of noise affecting all the $L = 8M$ watermark coefficients was found to be correlated with the original LP coefficients.

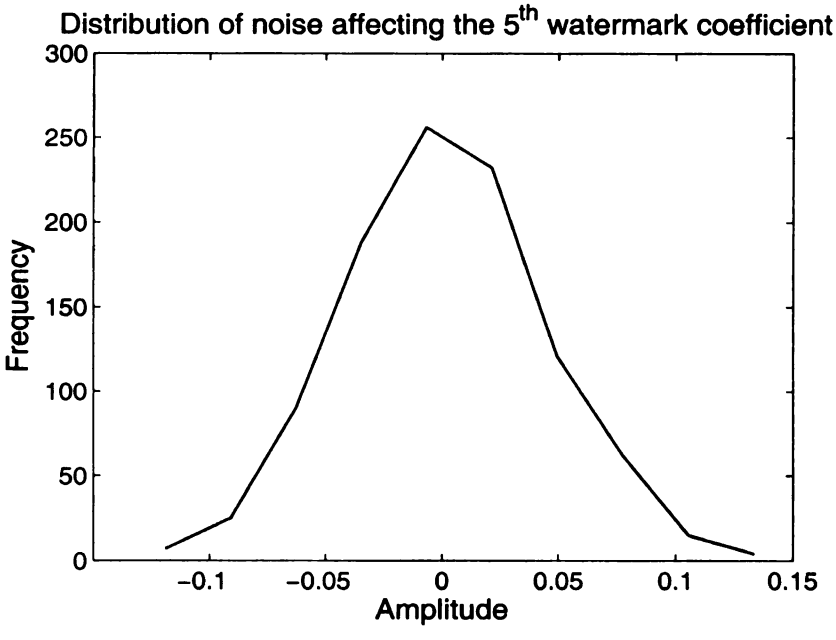
A solution to this problem is to normalize the watermark coefficients before adding them to the original LP coefficients. That is, instead of directly adding the watermark vector to the original LP coefficients ($\tilde{\mathbf{a}} = \mathbf{a} + \omega$), we obtain the modified LP coefficients as,

$$\tilde{a}_i = a_i + \omega_i |a_i|. \tag{3.15}$$

From the estimate of the modified LP coefficients, the watermark vector



(a)



(b)

Figure 3.1: Typical noise distribution in the LP domain for any coefficient. For Fig. 3.1(a) 15 dB white noise was added in time domain to the stegosignal, and for Fig. 3.1(b) 15 dB colored noise was added to the stegosignal.

is obtained as,

$$\hat{\omega}_i = \frac{\hat{\tilde{a}}_i - a_i}{|a_i|} \tag{3.16}$$

with $\hat{\tilde{\mathbf{a}}} = \{\hat{\tilde{a}}_i\}_{i=1}^{M}$, as defined in Table 3.2. However, when $|a_i| \ll 1$, the recovery of watermark coefficients magnifies the noise variance in the LP domain. To avoid this, watermark coefficients are normalized before embedding, but only if $|a_i| \geq 1$. For the experiments presented in the rest of the chapter, the watermark embedding and recovery involves this "selective normalization." Accordingly, Step 3 of Table 3.1 is carried out using the following rule in the present algorithm:

$$\tilde{a}_i = \begin{cases} a_i + \omega_i |a_i|, & \text{if } |a_i| \geq 1 \\ a_i + \omega_i, & \text{otherwise} \end{cases}.$$

The final step in the recovery algorithm (Table 3.2) involves the following equation:

$$\hat{\omega}_i = \begin{cases} (\hat{\tilde{a}}_i - a_i)/(|a_i|), & \text{if } |a_i| \geq 1 \\ \hat{\tilde{a}}_i - a_i, & \text{otherwise} \end{cases}.$$

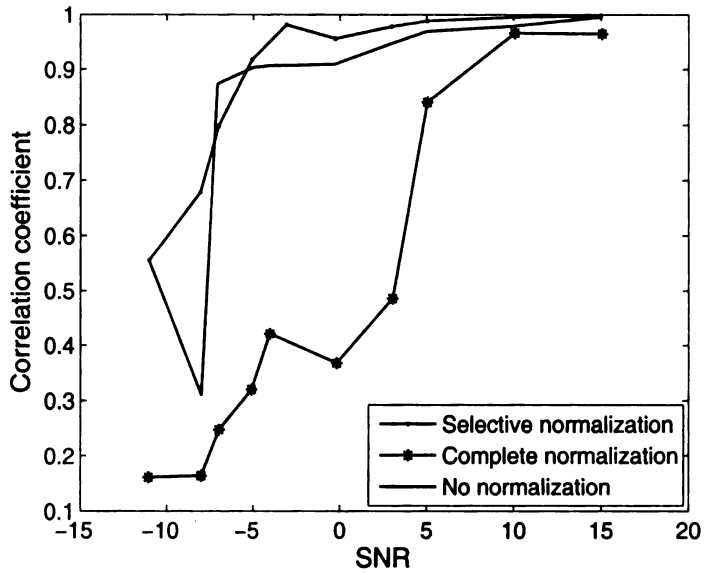
In Table 3.3, μ and σ^2 are the parameter noise mean and variance, and $c_{\mathbf{ra}}(0)$ is the cross-correlation between the recovered vector and the original LP coefficients. The values of μ, σ^2, $c_{\mathbf{ra}}(0)$ were determined by conducting 1000 experiments, involving a stegosignal of

Table 3.3: Effect of selective normalization

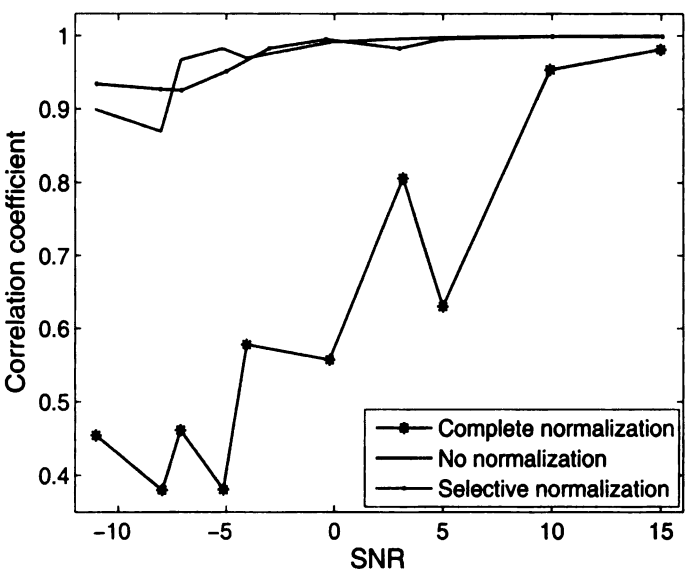
Noise	SNR (dB)	Normal -ization	μ	σ^2	$c_{\mathbf{ra}}(0)$
White	10	no	2.849×10^{-4}	0.0517	-0.0059
White	10	complete	-0.0152	4.6477	-0.0051
White	10	selective	-6.2×10^{-5}	0.1099	7.645×10^{-4}
Color	15	no	2.3438×10^{-4}	0.0049	0.0328
Color	15	complete	0.0139	0.7518	-0.0094
Color	15	selective	-1.162×10^{-4}	0.0071	0.0023

1 s duration. The stegosignal was watermarked at CWR_{seg} of 7 dB using a watermark message consisting of eight orthogonal vectors, each vector embedded into 0.125 seconds (2000 samples). Selective normalization of watermark coefficients significantly reduces the correlation between noise in LP domain and the LP coefficients, especially when the stegosignal is subjected to colored noise in the time domain (see Table 3.3). Moreover, as the noise variance in the LP domain is reduced, selective normalization improves the cross-correlation between the original and recovered watermarks compared to the complete normalization case. Figures 3.2(a) and (b) also show an improvement in the correlation coefficient values when selective normalization is used.

The watermark detection process is treated as a binary decision problem in the presence of additive noise. Preliminary watermark de-



(a)



(b)

Figure 3.2: Effect of complete normalization, selective normalization, and no normalization of watermark coefficients on the correlation coefficient between original and recovered watermarks. In 3.2(a) the stegosignal was distorted by white noise in the time domain, and in 3.2(b) colored noise was added to the stegosignal.

tection experiments are used to set the hypotheses,

$$H_0: \; r_i = v_i, \qquad i = 1, 2, ..., L$$
$$H_1: \; r_i = \omega_i + v_i, \qquad i = 1, 2, ..., L$$

where $\{r_i\}_{i=1}^{L}$ is the set of elements in the observation vector. The null hypothesis is that no watermark is present and only noise is transmitted $\{v_i\}_{i=1}^{L}$, while under H_1, both watermark $\{\omega_i\}_{i=1}^{L}$ and noise samples $\{v_i\}_{i=1}^{L}$ are present in additive combination. Due to selective normalization of watermark coefficients, noise in the LP domain, v_i is distributed as $\mathcal{N}(0, \sigma^2)$, when noise $\{\zeta_i\}_{i=1}^{N}$ is added to the stegosignal in the time domain such that the SNR is $S_1 = 10 \log_{10} \left[(\sum_{n=1}^{N} \tilde{y}_n^2)/(\sum_{n=1}^{N} \zeta_n^2) \right]$. For this watermark detection problem, the expressions for false-alarm, detection and missed-detection rates are well-known and are given by (e.g., [40]),

$$P_F = 0.5 \left[\mathbf{erfc} \left(\frac{\ln \tau + \frac{1}{2\sigma^2} \sum_{i=1}^{L} \omega_i^2}{\sqrt{2}\bar{\sigma}} \right) \right] \tag{3.17}$$

$$P_D = 0.5 \left[\mathbf{erfc} \left(\frac{\ln \tau + \frac{1}{2\sigma^2} \sum_{i=1}^{L} \omega_i^2 - \bar{\mu_1}}{\sqrt{2}\bar{\sigma}} \right) \right] \tag{3.18}$$

$$P_M = 1 - P_D \tag{3.19}$$

Here, $\bar{\mu_1} = (2\sigma^2)^{-1} \sum_{i=1}^{L} \omega_i^2$, $\bar{\sigma} = (2\sigma^2)^{-1} \sqrt{\sum_{i=1}^{L} \omega_i^2}$ and τ is the detection threshold. Let $\tau'' = \sigma^2 \ln \tau + 0.5 \sum_{i=1}^{L} \omega_i^2$ then, the decision rule

is

$$\sum_{i=1}^{L} r_i \omega_i \underset{H_0}{\overset{H_1}{\gtrless}} \tau''. \tag{3.20}$$

In a practical implementation, the threshold τ'', corresponding to an SNR of S_1, can be adjusted further if the actual SNR in the time domain is determined. As an example, if the SNR were found to be $S_2 = 10\log_{10}\left[(\sum_{n=1}^{N}\tilde{y}_n^2)/(\sum_{n=1}^{N}\check{\zeta}_n^2)\right]$ (assuming zero-mean noise), the threshold τ'' is altered by multiplying σ^2 with the adjustment factor $1/\beta$, where $\beta = 10^{(S_1-S_2)/10}$.

The SNR in the parameter domain is defined as, $d^2 = (\bar{\mu_1}/\bar{\sigma})^2$ [40]. In the present case, $d = \sqrt{\bar{\mu_1}}$. Hence embedded marks of greater energy will result in improved robustness, while noise of higher variance in the parametric domain will hinder watermark detection. The stegosignal was subjected to additive white and colored noise, resulting in different SNRs in the time and parameter domains. In each case, experiments were repeated 1000 times in order to estimate the mean and variance of the Gaussian noise affecting each watermark coefficient. Receiver operating characteristics (ROC) were determined using equations (3.17) and (3.18). It is observed in Table 3.4 that very low false-alarm rates

Table 3.4: Estimates of SNR, d^2, P_D and P_F

Noise	SNR (dB)	d^2	P_D	P_F	τ''
White	15	696.95	0.99999	4.37×10^{-114}	6.8699
White	10	72.79	0.99994	1.37×10^{-6}	4.3960
Colored	7	167.29	0.99999	1.20×10^{-18}	5.4038
White	3	14.45	0.9987	0.215	1.6610
White	1	9.54	0.99715	0.37304	0.8388

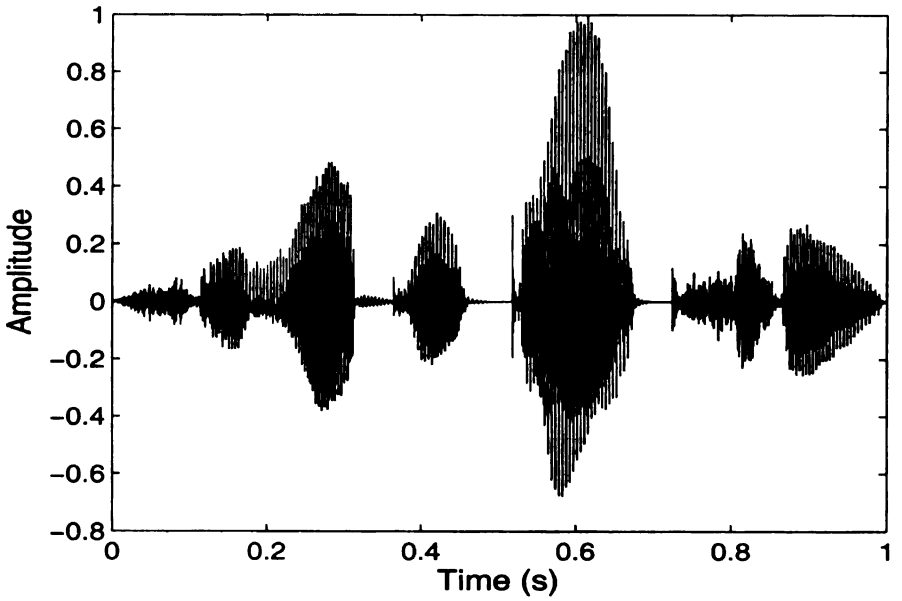
can be achieved using parametric watermaking with selective normalization. For example, when 10 dB white noise is added to the stegosignal, for a threshold $\tau'' = 4.3960$, a $P_D = 0.99999$ and a false-alarm rate $P_F = 1.37 \times 10^{-6}$ is obtained. Experiments were performed for time domain SNRs of 1 dB and 3 dB and P_F was found to be 0.14 and 0.0033 respectively, an improvement over the results in Table 3.4. It should be noted that for time domain SNRs below 10 dB, the resulting stegosignals are degraded to the point of being unuseful as surrogates for the coversignal. Comparing the SNR in the time and parameter domains it can be observed from Table 3.4 that parametric watermarking significantly boosts the SNR. The resulting parameter noise gain suppression contributes to improved watermark detection.

3.2 Experiments and discussion

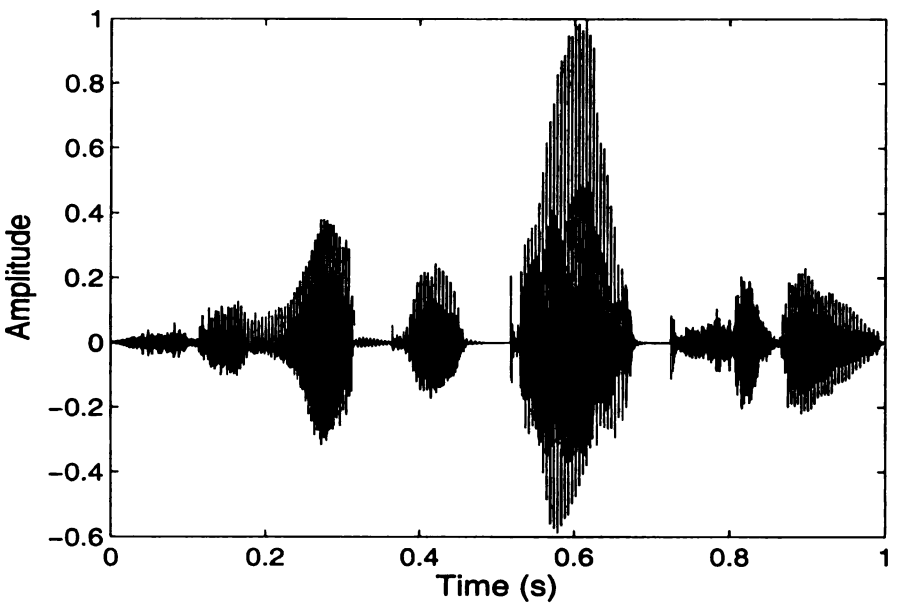
3.2.1 Introduction

Robustness refers to the ability of the watermark to tolerate distortion from any source to the extent that the quality of the coversignal is not affected beyond a set fidelity standard, or that the watermark detection and recovery processes are not hindered. Experiments performed to investigate the perceptual and robustness aspects of LP parametric watermarking are presented in this section. Some of the factors affecting the robustness of the present technique include the length of the speech frame to be watermarked, the choice of watermark sequence, the relative energy of the watermark, and the temporal locations and durations of the watermarks in the stegosignal. In broader terms, watermark robustness also depends on the watermark embedding, recovery, and detection algorithms.

For the experiments below, speech was watermarked using both LP based parametric and SS watermarking algorithms. Both LP and SS watermarking algorithms involve private decoding. In the experiments presented below, the coversignal [shown in Fig. 3.3(a)] consists of 1 s of speech from the TIMIT database [41], sampled at 16 kHz. The sentence "She had your dark suit in greasy wash water all year." is uttered by a female talker. For the robustness experiments, parametric



(a)



(b)

Figure 3.3: Plots of (a) coversignal and (b) stegosignal at CWR_{seg} of 7.715 dB. The coversignal and the stegosignal are of 1 s duration and sampled at 16 kHz. The speech is divided into frames of 2000 samples and a watermark vector is embedded into each of the eight frames.

watermarking is implemented at CWR_{seg}'s of 7.715 dB and 10.68 dB. SS watermarking was implemented at CWR_{seg}'s of 7.715 dB, 10.68, 27 dB, and 30 dB. This is explained further in Section 3.2.2. The sample correlation coefficient is used as the measure of similarity between original and recovered watermark vectors for both parametric and SS watermarking techniques. The correlation coefficient between two random variables ω and r is given by

$$\frac{\mathbf{c}_{\omega r}(0) - E(\omega)E(r)}{\sigma_\omega \sigma_r}, \tag{3.21}$$

where $\mathbf{c}_{\omega r}$ is the cross-correlation between ω and r, $E(\omega)$ and $E(r)$ are the expected values of ω and r, and σ_ω^2 and σ_r^2 are the variances of ω and r, respectively. For sample correlation coefficient, the expected values of ω and r are replaced by the samples means $m_\omega = \frac{1}{L}\sum_{i=1}^{L}\omega_i$ and $m_r = \frac{1}{L}\sum_{i=1}^{L} r_i$. And the variances, σ_ω^2 and σ_r^2, are replaced by sample variances of ω and r given by $var_\omega = \frac{1}{L}\sum_{i=1}^{L}(\omega_i - m_\omega)^2$ and $var_r = \frac{1}{L}\sum_{i=1}^{L}(r_i - m_r)^2$, respectively. The sample cross-correlation between ω and r at lag 0 is $\frac{1}{L}\sum_{i=1}^{L}\omega_i r_i$. Since the watermark vectors are mutually orthogonal, the correlation coefficient between distinct watermark vectors is 0.

Bit error rate is another commonly used performance measure of similarity between the original and recovered watermarks. The bit error rate (BER) is defined as the ratio of number of bit errors to total

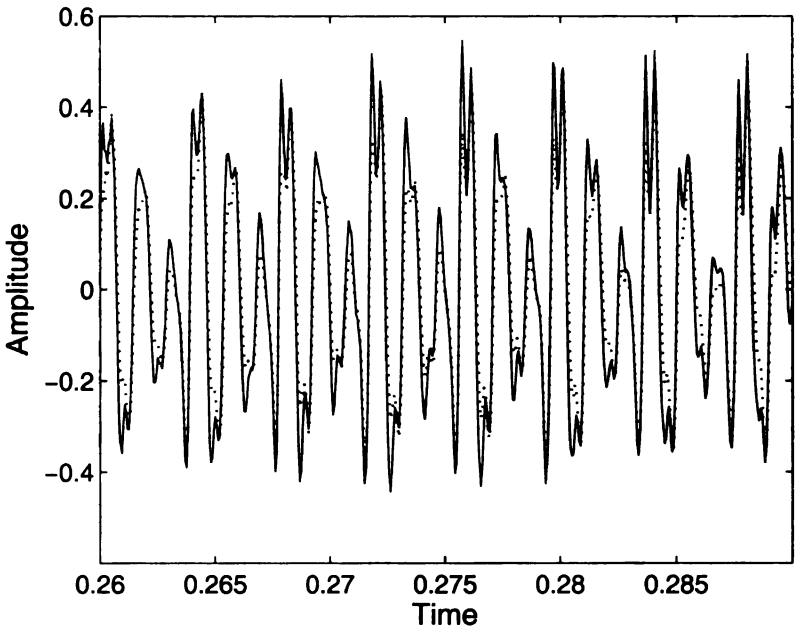


Figure 3.4: Segments of cover (dotted line) and stegosignals (continuous line) of 480 samples or of 0.03 ms duration and a CWR_{seg} of 7.715 dB. The cover and stegosignals used in the robustness experiments are of 1 s duration and sampled at 16 kHz. The speech is divided into frames of 2000 samples and a watermark vector is embedded into each of the eight frames.

number of bits transmitted. In this work, it is more relevant to use correlation coefficient than BER because it is more important to characterize the performance based on the detection and recovery of the entire watermark vector rather than the individual bits or watermark vector elements. The probability of signal or watermark vector error is also a useful performance measure. The relation between correlation coefficient, probability of signal error and BER can be found in [42]. In a practical implementation, a recovered vector, possibly containing the watermark, is first sent to the detector of Section 3.1.5, which is governed by the decision rule in equation (3.20).

For LP parametric watermarking, the speech was divided into eight frames of 2000 samples each or 0.125 seconds duration. The watermarks were comprised of (eight) non-binary orthogonal vectors of length eight. In each of the speech frames, a length eight watermark vector was embedded into to the coefficients of an eighth-order LP model, resulting in a bit rate of 24 bits per second. For the parametric watermarking experiments presented in this section, the watermark embedding and recovery involved selective normalization. The resulting stegosignal [Fig. 3.3(b)] was subjected to various attacks discussed below.

For the SS algorithm, the stegosignal $\{\breve{y}_j\}_{j=1}^{N}$ was obtained by adding the watermark sequence $\{g_i\}_{i=1}^{1000}$ to the 1000 largest DCT coef-

ficients of the coversignal of 1 s duration.

$$\breve{Y}_i = Y_i(1 + \lambda g_i),$$

where every g_i is independently drawn from $\mathcal{N}(0, 1)$, and Y_i and $\breve{Y}_i$ are the i^{th} largest DCT coefficients of the coversignal and stegosignal, respectively. The λ parameter is adjusted to obtain a desired CWR_{seg}.

3.2.2 Subjective perceptual tests

Although CWR_{seg} is used as the objective measure of fidelity, listening tests were also performed to compare the watermarked speech fidelity. Speech was watermarked using both parametric and SS algorithms for CWR_{seg} ranging from 1 dB to 40 dB.

For the robustness experiments discussed in the following section, two implementations of *LP parametric* watermarking at CWR_{seg} of 7.715 dB and 10.68 dB were used. Parameter-embedded watermarks were inaudible at these or higher CWR_{seg} [30]. Different CWR_{seg}values can be selected depending on the fidelity constraint for a given application. For performance comparison, implementations of SS watermarking at 7.715 dB and 10.68 dB were also used. Additionally, listening tests were performed, to subjectively identify CWR_{seg}'s of SS-watermarked stegosignals, whose fidelity was comparable to the 7.715 dB and 10.68 dB implementations of parametric watermarking. This

was imperative, as an objective measure such as CWR_{seg}, although an improvement over CWR, does not satisfactorily quantify all the perceptual aspects of fidelity. Five subjects were asked to select the SS watermarking implementations that sounded most similar to the 7.715 dB and the 10.68 dB implementations of LP parametric watermarking, from a set of stegosignals with CWR_{seg}'s ranging from 1 dB to 40 dB. The sounds files used in the listening tests are available at the website [30]. Based on the subjective tests it was concluded that the 7.715 dB implementation of LP parametric watermarking was perceptually similar to the 27 dB implementation of SS watermarking, and the fidelity of 10.68 dB implementation of LP parametric watermarking was comparable with 30 dB implementation of SS watermarking. This, in itself, is significant because it demonstrates the fidelity benefits that can be achieved through parametric watermarking.

3.2.3 Watermark robustness

In this section, we analyze the robustness to common attacks of watermarks inserted by LP based parametric watermarking. The stegosignals used in these experiments were obtained by embedding watermarks through direct manipulation of the LP coefficients. The SS watermarking algorithm for multimedia signals [10] was used to benchmark performance.

For meaningful analysis of detection performance, it is necessary to consider stationary segments of the coversignal and the stegosignal. That is, segments of y_n, w_n, and, hence, $\tilde{y}_n$, are assumed to be partial realizations of wide-sense stationary (WSS) and ergodic random processes. Generally, speech sequences can be considered stationary across frames of duration 20 ms. However, the robustness experiments presented below are based on speech frames of longer duration, typically, 125 ms, in order to balance the conflicting requirements of stationarity and longer frame lengths for the LSE estimation and stegosignal fidelity. Hence, one important observation to be made based on the experimental results is the effect of non-stationarity on watermark robustness.

Robustness to additive white noise attack

Let $\{\eta_n\}_{n=1}^N$ be a partial realization of a zero mean, uncorrelated noise process which is added to the stegosignal samples $\{\tilde{y}_n\}_{n=1}^N$. Let the corrupted stegosignal be denoted $\{\tilde{y}_n^\eta\}_{n=1}^N$. In this case, the "output" signal used in the LSE problem (equation (3.5) for $n \in [1, N]$) will be likewise corrupted. That is, the clean signal d_n is replaced by, say,

$$d_n^\eta = \tilde{y}_n^\eta - \xi_n = d_n + \eta_n, \; n = 1, 2, \ldots, N. \tag{3.22}$$

Accordingly, the cross-correlation vector $\mathbf{c}_{yd^\eta}$ [i.e., right side of normal

equations (3.6)], but *only* this vector, is affected by the attack. The LSE solution is

$$\tilde{\mathbf{a}}^{\eta} = \mathbf{C}_y^{-1}\mathbf{c}_{yd^{\eta}} = (\mathbf{Y}_N\mathbf{Y}_N^T)^{-1}\mathbf{Y}_N\mathbf{d}_N^{\eta} \tag{3.23}$$

where, $\mathbf{d}_N^{\eta} = \begin{bmatrix} d_N^{\eta} & d_{N-1}^{\eta} & \cdots & d_1^{\eta} \end{bmatrix}^T \in \mathbb{R}^N$. Equation (3.23) can be expressed as,

$$\tilde{\mathbf{a}}^{\eta} = \mathbf{C}_y^{-1}\mathbf{c}_{yd} + \mathbf{C}_y^{-1}\mathbf{c}_{y\eta} = \tilde{\mathbf{a}} + \mathbf{C}_y^{-1}\mathbf{c}_{y\eta}. \tag{3.24}$$

The i^{th} value of the cross-correlation term $\mathbf{c}_{y\eta}$ is given by $\mathbf{c}_{y\eta}(i) = \sum_{n=1}^{N} y_{n-i}d_n$. Since the noise is uncorrelated, $\mathbf{c}_{y\eta}$ asymptotically as $N \to \infty$ approaches the zero vector $\mathbf{0}$. Hence the "corrupted" cross-correlation, $\mathbf{c}_{yd^{\eta}}$, approaches $\mathbf{c}_{yd}$ for large N. The watermark is therefore asymptotically immune to the white noise attack. In the presence of white noise, $\tilde{\mathbf{a}}^{\eta}$ is an unbiased and consistent estimator of $\tilde{\mathbf{a}}$ for all N.

To verify the analysis, experiments were performed in which white Gaussian noise resulting in different SNRs was added to speech watermarked by both LP and SS algorithms. The correlation coefficients between the original and recovered watermarks from all eight stegosignal frames were determined and averaged. It is seen in Fig. 3.5 that, at any SNR, LP parametric watermarking at 7.715 dB and 10.68 dB

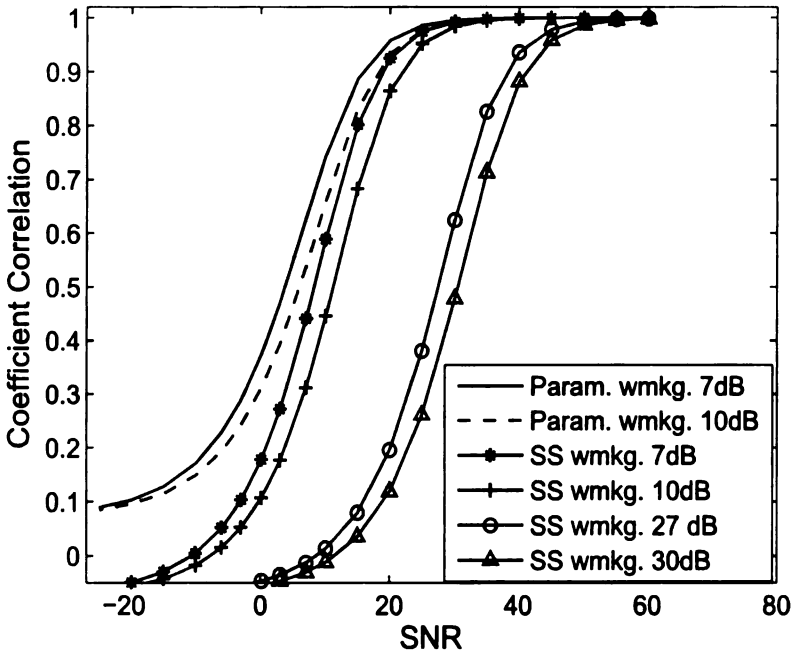


Figure 3.5: Watermark robustness to white noise attack. Performance of parametric watermarking at CWR_{seg}'s of 7.715 dB and 10.68 dB is compared with that of SS watermarking at 7.715 dB, 10.68 dB, 27 dB and 30 dB CWR_{seg}.

CWR_{scg} results in higher correlation between original and recovered watermarks compared to SS watermarking at CWR_{seg}s of 7.715 dB, 10.68 dB, 27 dB or 30 dB. This improvement in the correlation coefficient values, and hence robustness, is mainly due to the LSE-based recovery algorithm. This level of robustness to white noise attack is sufficient for a wide-range of watermarking applications, as the stegosignal is highly noisy below an SNR of 15 dB (for details see [30]). The non-stationarity of the 2000-sample watermarked speech frame can be ignored for practical applications of parametric watermarking. Also, as expected, watermark robustness to attack increases as the CWR_{seg} is decreased, since there is greater watermark energy in the same coversignal.

Robustness to colored noise attack

In the next set of experiments, the stegosignal segment was distorted by the addition of a colored noise process, $\{\gamma_n\}_{n=1}^{N}$. Colored noise was generated by filtering a white noise process using a 11^{th}order FIR lowpass filter with a cut-off frequency of 0.4 (normalized) or 6400 Hz. The distorted stegosignal frame is denoted $\{\tilde{y}_n^{\gamma}\}_{n=1}^{N}$. As in the white noise case, the "output" signal in the watermark recovery process is corrupted. Instead of d_n, we have access to

$$d_n^{\gamma} = \tilde{y}_n^{\gamma} - e[n] = d_n + \gamma_n, \quad n = 1, 2, ..., N. \tag{3.25}$$

Consequently, the cross-correlation vector in the normal equations is altered by the attack. Because of the correlation in the noise, $\mathbf{c}_{yd^\gamma}$ no longer approaches $\mathbf{c}_{yd}$ asymptotically. Depending on the relative magnitudes of the cross-correlation elements in $\mathbf{c}_{yd^\gamma}$, the LSE estimation of the perturbed coefficients, and hence the watermark, will be affected. The solution to this problem is a prewhitening procedure.

In the presence of colored noise, the LSE estimation problem is represented by the following equation,

$$\mathbf{Y}_N^T \tilde{\mathbf{a}}^\gamma = \mathbf{d}_N + \gamma_N. \tag{3.26}$$

in which, $\gamma_N = \begin{bmatrix} \gamma_1 & \gamma_2 & \cdots & \gamma_N \end{bmatrix}^T$ and all other quantities are defined above. Pre-multiplying both sides of equation (3.26) by the inverse covariance matrix of the colored noise, $\mathbf{C}_\gamma^{-1}$, and rearranging the terms, results in

$$\tilde{\mathbf{a}}^\gamma = (\mathbf{Y}_N \mathbf{C}_\gamma^{-1} \mathbf{Y}_N^T)^{-1} \mathbf{Y}_N \mathbf{C}_\gamma^{-1} \mathbf{d}_N^\gamma = (\mathbf{Y}_N \mathbf{C}_\gamma^{-1} \mathbf{Y}_N^T)^{-1} \mathbf{Y}_N \mathbf{C}_\gamma^{-1} \mathbf{d}_N. \tag{3.27}$$

Thus, the estimation of the perturbed LP coefficients is the solution to (3.27) with $\mathbf{C}_y$ replaced by $\mathbf{C}_y^\gamma = (\mathbf{Y}_N \mathbf{C}_\gamma^{-1} \mathbf{Y}_N)$ and $\mathbf{c}_{yd}$ replaced by $\mathbf{c}_{yd}^\gamma = \mathbf{Y}_N \mathbf{C}_\gamma^{-1} \mathbf{d}_N$. Whitening requires knowledge of noise correlation properties which are readily determined in the present application.

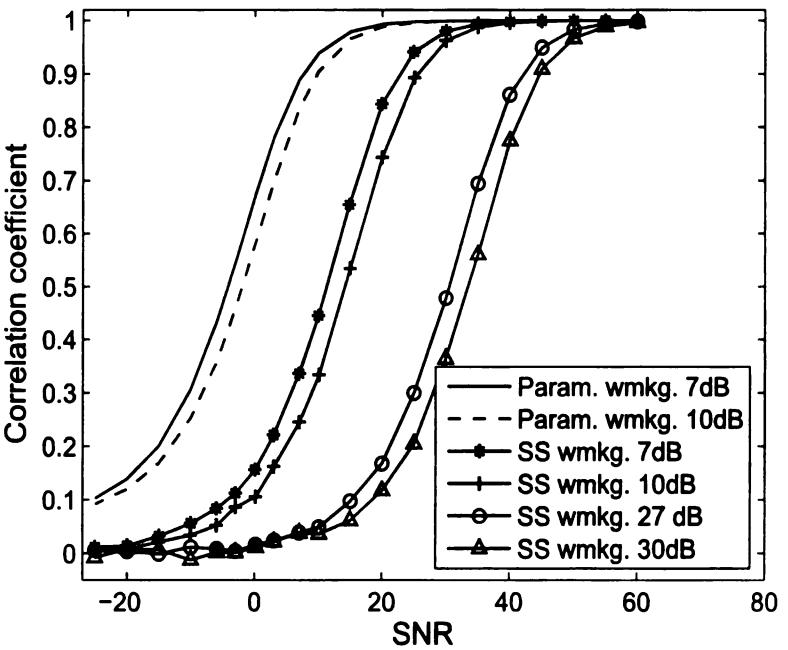


Figure 3.6: Watermark robustness to colored noise attack. Colored noise was generated by lowpass filtering white noise.

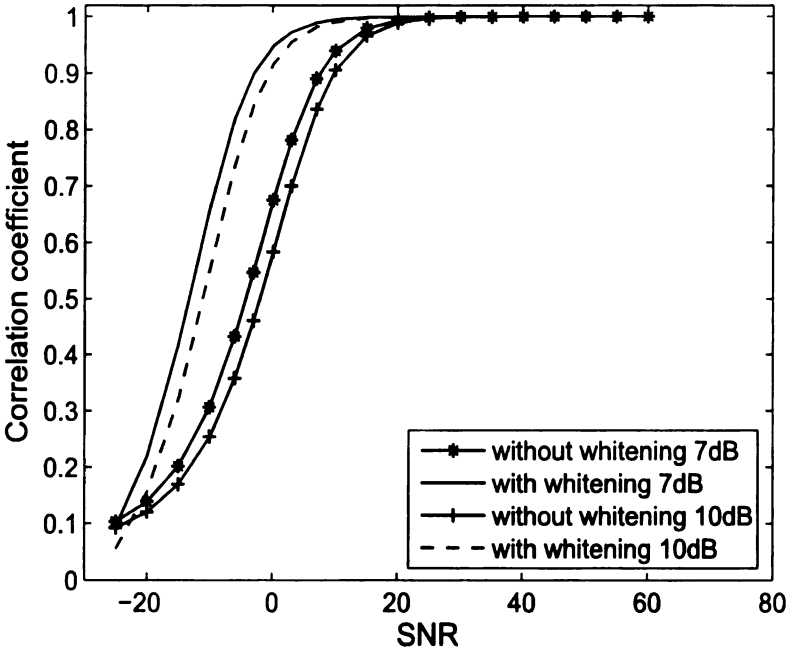


Figure 3.7: Improvement in watermark robustness to colored noise attack due to whitening transformation.

The effect of colored noise on watermark robustness is represented in Fig. 3.6. It is observed that LP parametric watermarking is fairly robust to colored noise, even in the *absence* of a prewhitening operation during the recovery process. In Fig. 3.6, the differences in performance between parametric and SS watermarking algorithms at 7.715 and 10.68 dB is even greater than in case of white noise (Fig. 3.5). An improvement in watermark robustness to colored noise attack is observed in Fig. 3.7, where the watermark recovery process involves prewhitening. In fact, LP parametric watermarking with prewhitening at 10.68 dB results in better robustness than at 7.715 dB without whitening, even though the latter outperforms SS at 7.715 dB.

Robustness to filtering

Let $\{\tilde{y}_n^f\}_{n=1}^N$ be the result of filtering the stegosignal. At time n,

$$\tilde{y}_n^f = \tilde{y}_n * h_n = y_n * h_n + w_n * h_n, \tag{3.28}$$

where $\{h_n\}$ is the impulse response of the filter, $*$ denotes linear convolution, and where we have continued to denote the watermark signal $w_n = \tilde{y}_n - y_n$.

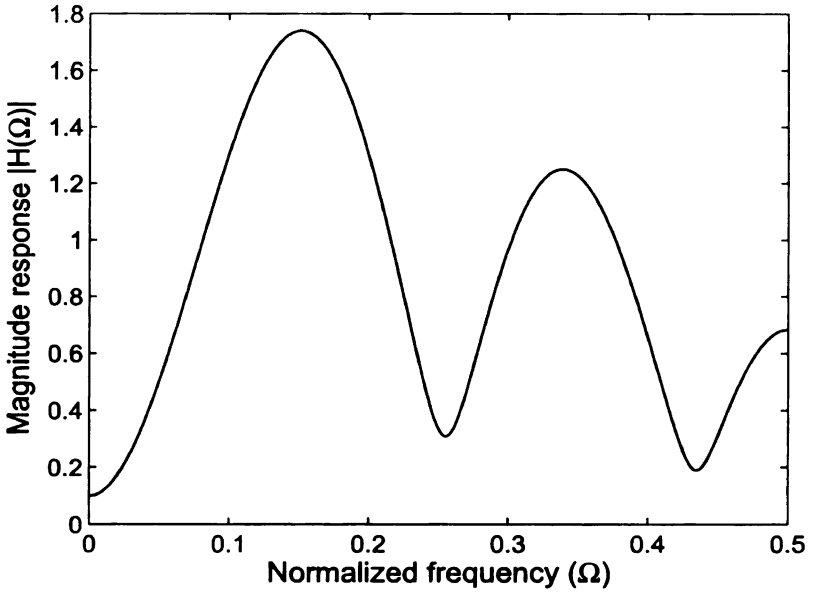
In the first analysis, it seems very reasonable that an ideal attack would be designed to result in $\tilde{y}_n^f \approx y_n$. This indicates that the ideal attack filter will maximize (in some sense) the contribution of the first

term in the sum in (3.28), and minimize the second – similar to any optimal filter design to remove noise.[1] On the other hand, (3.28) reveals that good watermark design requires that the watermark signal be as spectrally similar to the coversignal as possible, so that any attack on the watermark will also degrade the coversignal component, thereby degrading fidelity. Since the effectiveness of an attack is constrained by the perceptual distortion of the stegosignal, for robustness to filtering attacks, it is sufficient that most of the watermark information be present in perceptually significant components of the coversignal [10]. In general, speech signals have most of the perceptually significant components in the low frequency spectrum, and hence watermark signals with low frequency spectra are most likely to survive a filtering attack - assuming that the attacker uses a rational approach which preserves fidelity.

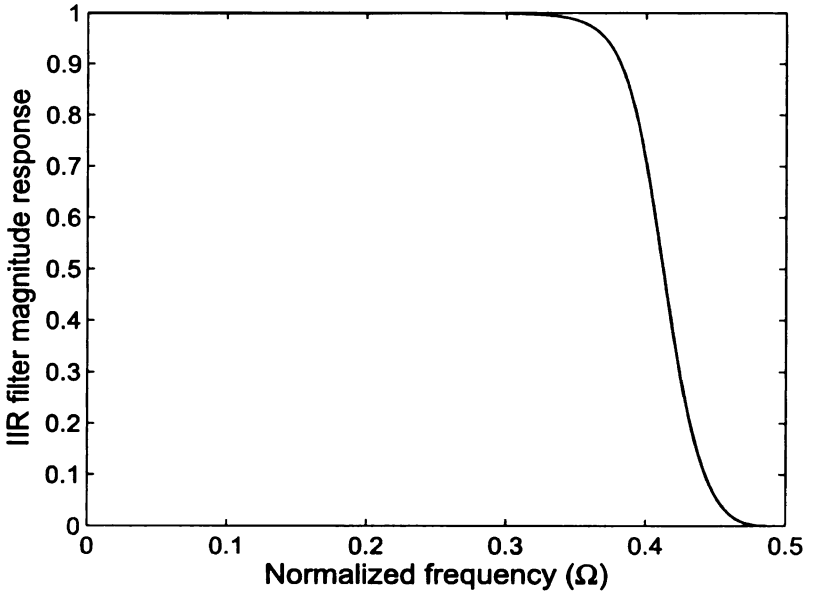
Watermark robustness to a 4^{th}-order butterworth lowpass filter for a range of cut-off frequencies is shown in Fig. 3.9. Since the watermark vector can be interpreted as coefficients of an FIR filter $\{\omega_i\}_{i=1}^{M} = \{\varpi[i]\}_{i=1}^{M}$, the magnitude response of this FIR filter ($|\mathcal{W}(\Omega)|$) is as shown in Fig. 3.8(a), while the magnitude response of the attack filter is shown in Fig. 3.8(b).

Watermark robustness to filtering depends on the magnitude spec-

[1]Since the attacker does not have access to the watermark signal w_n, truly optimal design - from the attacker's point of view - is not possible.



(a)



(b)

Figure 3.8: Plots of (a) Magnitude spectrum of the watermark coefficients, and (b) Magnitude response of the attack filter at a normalized cut-off frequency of 0.4. A 4^{th}-order IIR Butterworth filter was used to test watermark robustness to lowpass filtering.

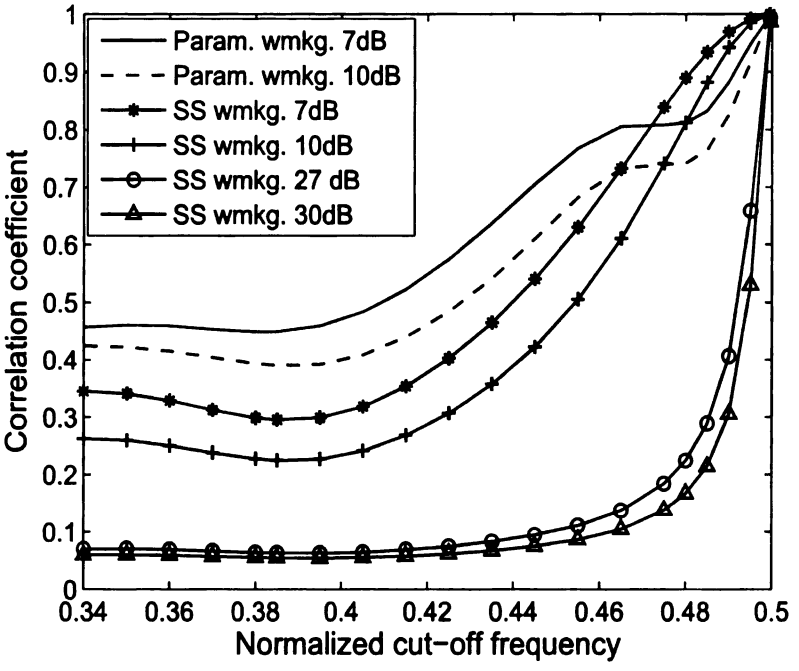


Figure 3.9: Robustness to lowpass filtering. A 4^{th}-order IIR butterworth filter was used to implement the lowpass filtering attack.

trum of the embedded watermarks. Low-frequency and mid-frequency watermark filters contribute to better robustness against lowpass filtering. Watermark robustness can be improved further through diversity, by repeatedly embedding watermark information [43]. Any highpass watermark filter $\{\varpi_{hp}[i]\}_{i=1}^{M}$ can be transformed into a lowpass watermark $\{\varpi_{lp}[i]\}_{i=1}^{M}$, using the relation [44]

$$\varpi_{hp}[i] = (-1)^i \varpi_{lp}[i]. \tag{3.29}$$

Robustness is improved by embedding the "same" watermark twice, in the original form $\{\omega_i\}_{i=1}^{M} = \{\varpi'[i]\}_{i=1}^{M}$ and in the frequency-translated form $\varpi[i] = (-1)^i \varpi'[i]$. The recovered watermark that has a higher correlation with the embedded watermark is used for watermark detection.

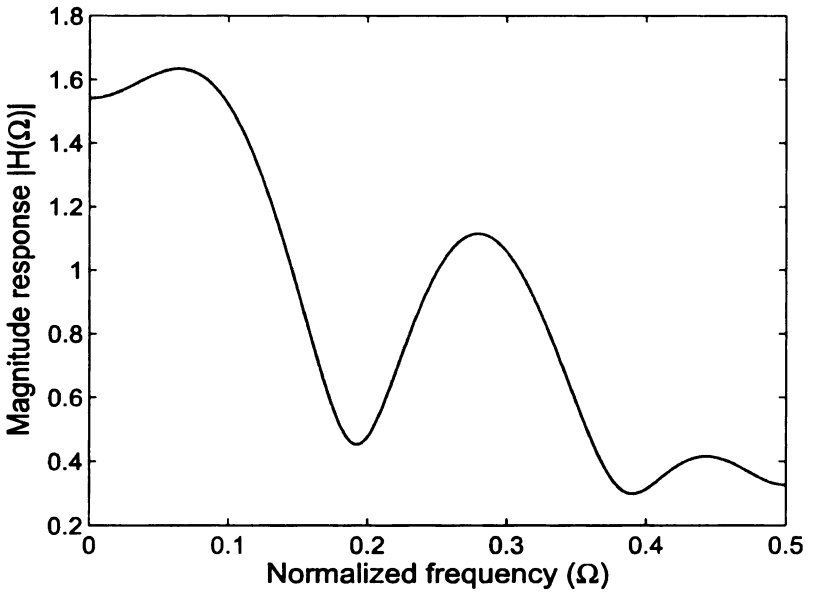
In order to illustrate this point, the coversignal was altered using the watermark whose magnitude spectrum is shown in Fig. 3.10(a), and its translated counterpart shown in Fig. 3.10(b). The resulting stegosignals were subjected to a highpass filtering attack using a 4^{th}-order butterworth filter. Figures 3.11(a) and 3.11(b) show that the transformed watermarked results in improved robustness to filtering. Highpass filters have a deleterious effect on speech quality. Even a cut-off frequency of 0.04 (normalized) or 640 Hz resulted in significant distortion of the stegosignals making them unusable for typical commer-

cial use, for example, and certainly for the digital library application addressed here.

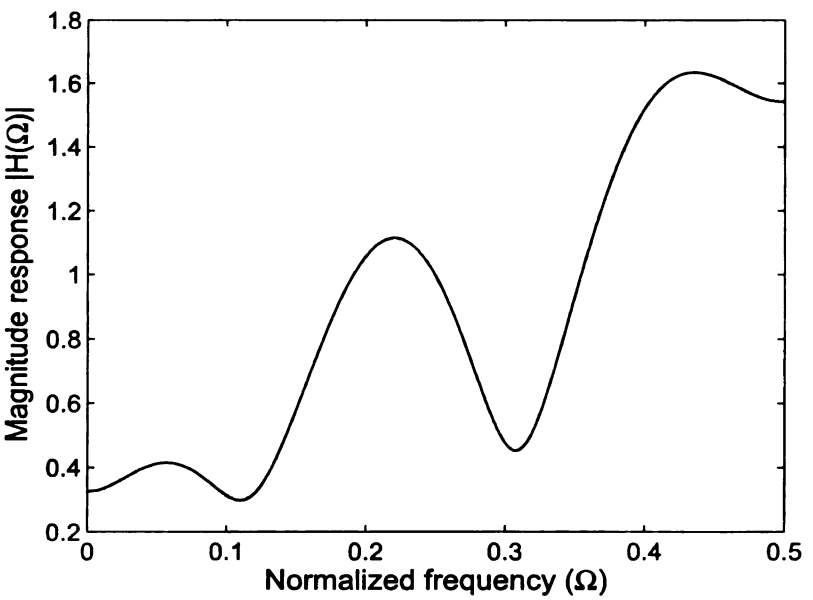
Robustness to cropping

In a *cropping* attack, arbitrary samples of the stegosignal are removed. Since the parametric modeling based watermarking involves an additive operation during the watermark embedding and recovery processes, cropping results in desynchronization of the coversignal and the stegosignal. However, as the present method is an informed watermarking technique, the algorithm described in [5] can be used for resynchronization of the cover and stegosignals.

In the present experiment, the stegosignal was subjected to a modified version of cropping, sometimes called the *jitter attack* [45]. In this modified implementation, random samples of the stegosignal were replaced by zeros. A specified percentage of samples from each frame of 2000 samples were randomly replaced by zeros. The fact that watermark information is spread-out in the stegosignal, while it is concentrated during the recovery process involving LSE, contributes to increased robustness of LP parametric watermarking to cropping as shown in Fig. 3.12.

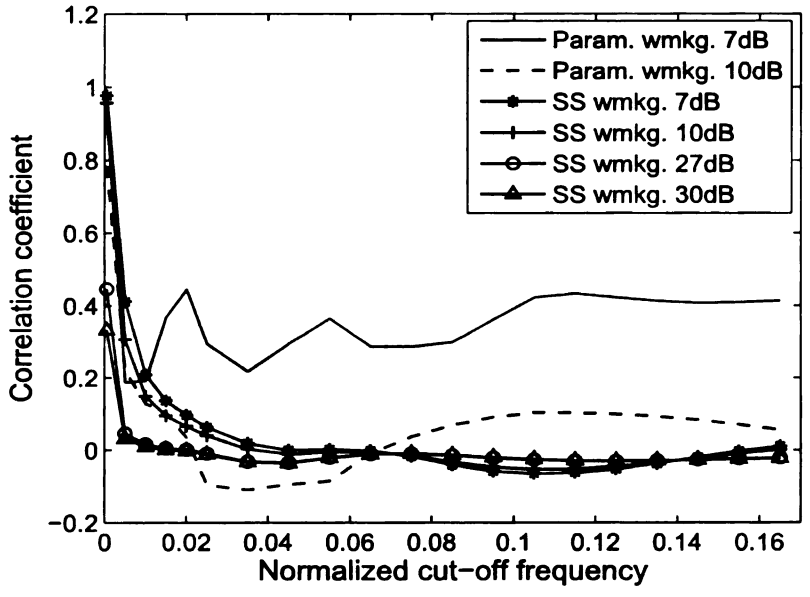


(a)

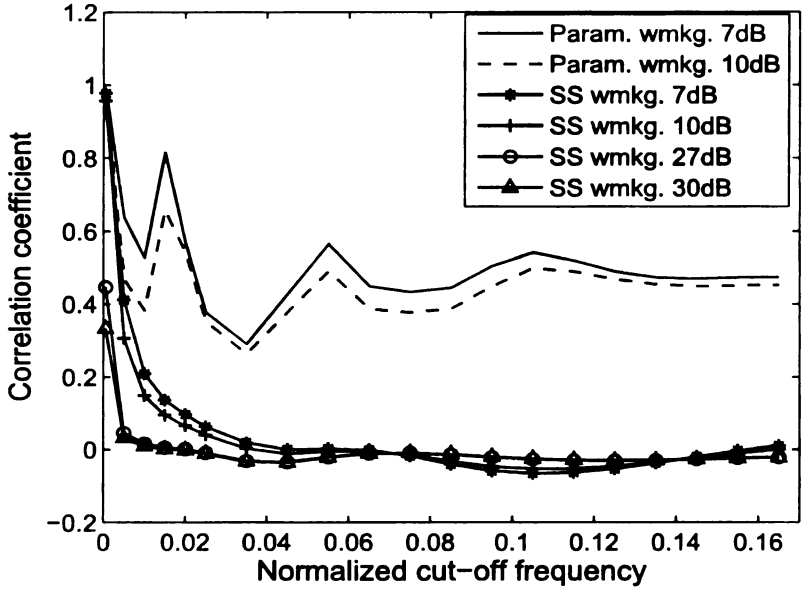


(b)

Figure 3.10: Plots of (a) Magnitude spectrum of the original watermark coefficients $h[n]$, and (b) Magnitude response of the transformed watermark coefficient, $(-1)^n h[n]$.



(a)



(b)

Figure 3.11: Robustness to 4^{th}-order butterworth highpass filter. In (a), the embedded watermark coefficients corresponded to a magnitude spectrum shown in Fig. 3.10(a), and in (b) the watermark coefficients were transformed using equation (3.29) and embedded.

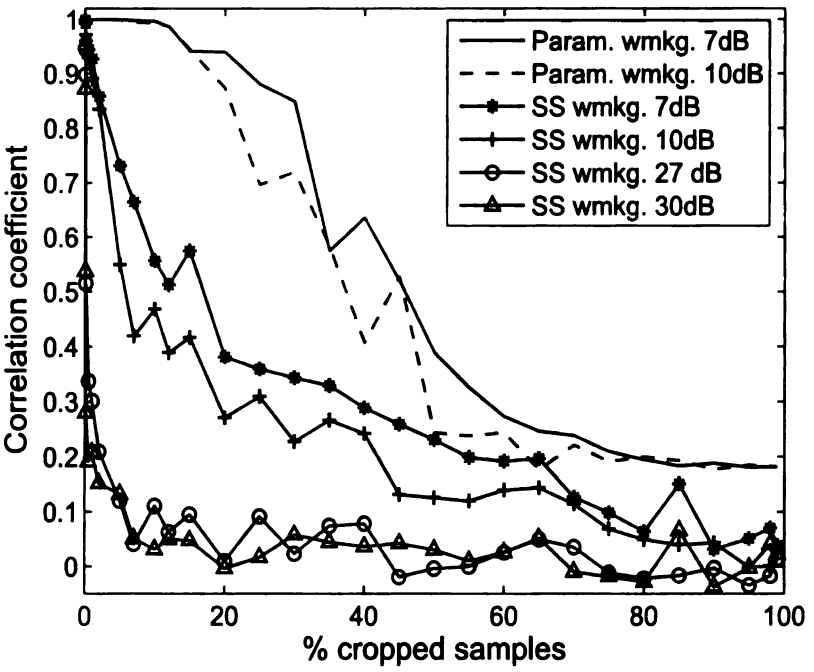


Figure 3.12: Robustness to cropping. Samples of the stegosignal were randomly cropped. Parameter-embedded watermarking results in improved robustness to cropping.

Table 3.5: Robustness to speech coding

Speech codec	Bit rate	Para wmkg 7 dB	Para wmkg 10 dB	SS wmkg 7 dB	SS wmkg wmkg 10 dB
	k bits/s	corr coef	corr coef	corr coef	corr coef
G.711	64	0.9990	0.9998	0.9985	0.9966
ADPCM	32	0.9889	0.9682	0.8207	0.7658
GSM	13.2	0.6584	0.5545	0.4095	0.3140
CELP	4.5	0.1488	0.1490	−0.0225	−0.0290
CELP	2.3	0.1269	0.1464	0.0209	0.0472
LPC10	2.4	0.1762	0.1762	−0.0184	−0.0470

Robustness to speech coding

Experiments were performed to study the effect of low-bit rate speech compression on LP based parametric watermarking. The stegosignal was compressed (coded), then decompressed (decoded), and the watermarks were recovered from the decompressed (decoded) signal. The correlation coefficient between the original and recovered watermarks for different speech codecs are tabulated in Table 3.5. The attacked stegosignals are available in the website [30]. The G.711 μ law, G.726 ADPCM, GSM (13.2 k bits/s), LPC10, and CELP (4.5 k bits/s) codecs were obtained from the website [46]. The G.711 speech codec software uses logarithmic pulse code modulation (PCM) and operates at sampling frequency of 8 KHz, with 8-bits per sample to compress and decompress speech. The G.726 uses adaptive differential PCM technique and is widely used in VoIP applications. The GSM full rate codec uses

an 8^{th} order linear prediction along with 13-bit uniform PCM. CELP and LPC10 codecs are also based on parametric models of speech. At 4.5k bits per second or less, the attacked stegosignals are intelligible, but are of low fidelity [30].

It is seen from Table 3.5 that the compression bit rate and CWR_{seg} are the main factors influencing watermark robustness. LP parametric watermarking outperforms SS watermarking in the presence of speech coding for all the bits rates tested. The performance of both SS and parametric watermarking is degraded significantly due to low bit rate CELP, GSM and LPC10 coding. However, the quality of the decompressed speech is also degraded considerably for CELP, GSM and LPC10 codecs [30]. Parameter-embedded watermarks are slightly more robust to LPC10 coding than CELP coding at 2.3 k bits/s. Although parametric watermarking involves perturbation of parameters of significance to speech coding, it performs better than SS watermarking in the presence of CELP, GSM and LPC10 codecs. This is because of the more speech-like rather than noise-like characteristic of the LP based watermark signal. At the same time, since LP analysis is performed over nonstationary segments of speech and synthesis is not involved in stegosignal reconstruction, robustness to a particular speech coder is not at the expense of the robustness to other codecs.

Chapter 4

LP Parametric Watermarking with a Fidelity Constraint

4.1 Introduction

The previous chapter described a speech watermarking algorithm wherein the LP parameters of the coversignal were modified by the addition of the watermark vector that was selected independently of the coversignal. The stegosignal was constructed using the correspondingly perturbed LP coefficients and the exact prediction residual, $\{\xi_n\}_{n=1}^N$, using the FIR filter $\tilde{y}_n = \sum_{i=1}^M \tilde{a}_i y_{n-i} + \xi_n = \tilde{\mathbf{a}}^T \mathbf{y}_n + \xi_n$. LP based parametric watermarking was found to be fairly robust against a wide variety of attacks such as addition of noise, MP3 compression, and cropping [7]. A main reason for good robustness is that the watermark signal is concentrated into a parametric representation during watermark embedding and recovery, while it is spread across the en-

tire work otherwise. Stegosignal fidelity and watermark robustness can be improved further if the embedded watermarks are obtained by integrating a fidelity constraint with the watermark embedding process and this led to SMF based parametric watermarking.

4.2 SMF parametric watermarking

SMF-based parametric watermarking subject to an ℓ_∞ fidelity constraint [29, 47] represents a step toward quantifying the relationship between the competing requirements of robustness and fidelity. The following general problem is addressed in this research:

CONSTRAINED WATERMARKING PROBLEM. *For coversignal frame* $\{y_n\}_{n=1}^N$ *generated according to model (3.2), find the set of watermarks, such that, for stegosignal frame* $\{y_n\}_{n=1}^N$ *generated according to (3.3), the following fidelity criterion is met,*

$$\|\mathbf{y} - \tilde{\mathbf{y}}\|_\infty < \gamma \tag{4.1}$$

in which $\mathbf{y}$ *and* $\tilde{\mathbf{y}}$ *are* N*-vectors with* n^{th} *elements* y_n *and* $\tilde{y}_n$*, respectively.*

In the present work, the determination of a watermark set guaran-

teed to satisfy a fidelity criterion is readily solved as an SMF problem (refer Section 2.2). First, let us subtract y_n from each side of equation (3.3), negate each side, then rearrange to obtain

$$y_n - \tilde{y}_n = (y_n - \xi_n) - \sum_{i=1}^{M} \tilde{a}_i y_{n-i} = (y_n - \xi_n) - \tilde{\mathbf{a}}^T \mathbf{y}_n. \tag{4.2}$$

Given a coversignal $\{\mathbf{y}_n \in \mathbb{R}^M\}_{n=1}^N$, a desired stegosignal $\{\tilde{y}_n \in \mathbb{R}\}_{n=1}^N$, and a maximum error tolerance γ, SMF [25] can be used to obtain the hyperellipsoidal membership set that tightly bounds the following feasibility set ($\mathcal{P}_N \subseteq \mathbb{R}^M$) at time N,

$$\mathcal{P}_N = \{\tilde{\mathbf{a}} \,|\, \|\mathbf{y} - \tilde{\mathbf{y}}\|_\infty < \gamma\}. \tag{4.3}$$

in which $\mathbf{y}$ is the N-vector with n^{th}element y_n, and $\tilde{\mathbf{y}}$ is the N-vector with n^{th}element $\tilde{\mathbf{a}}^T \mathbf{y}_n + \xi_n$.

The fidelity constraint can be generalized to allow for more "local" fidelity considerations in time as the signal properties change. A fidelity criterion takes the form of pointwise absolute error bounds, $\{\gamma_n\}_{n=1}^N$, on the difference between the stego- and coversignals: $|y_n - \tilde{y}_n| < \gamma_n$ for each $n \in [1, N]$. Upon defining the sequence

$$z_n = y_n - \xi_n, \quad n = 1, 2, \ldots, N, \tag{4.4}$$

(recall that $\{\xi_n\}$ is known) and the search for the constrained water-

mark parameters is reduced to a SMF problem as in (2.2). Applying SMF method to the estimation of $\tilde{\mathbf{a}}$ as in equation (4.2) yields hyper-ellipsoidal set of watermark (perturbed model parameter) candidates, $\boldsymbol{\mathcal{E}}_N$, guaranteed to contain and tightly bound the following exact set

$$\boldsymbol{\mathcal{P}}_N = \{\tilde{\mathbf{a}} \in \mathbb{R}^M \,|\, |z_n - \tilde{\mathbf{a}}^T \mathbf{y}_n| < \gamma_n, \;\; n \in [1, N]\}. \tag{4.5}$$

The fidelity constraint is a bound on $|w_n|$, where the watermark signal is given by $w_n = \tilde{y}_n - y_n$. The hyperellipsoidal set is,

$$\boldsymbol{\mathcal{E}}_N = \{\tilde{\mathbf{a}} | (\tilde{\mathbf{a}} - \mathbf{a}_c(N))^T \frac{\mathbf{C}_N}{\kappa_N} (\tilde{\mathbf{a}} - \mathbf{a}_c(N)) < 1\}, \tilde{\mathbf{a}} \in \Re^M \tag{4.6}$$

where $\mathbf{a}_c(N)$ is the center of $\boldsymbol{\mathcal{E}}_N$. $\mathbf{C}_N \in \mathbb{R}^{M \times M}$ is the covariance matrix and $\mathbf{C}_N = \mathbf{Y}_N \mathbf{Y}_N^T$ where $\mathbf{Y}_N = \begin{bmatrix} \mathbf{y}_N & \mathbf{y}_{N-1} & \cdots & \mathbf{y}_1 \end{bmatrix} \in \mathbb{R}^{M \times N}$. As shown in Table 2.2.2, κ_n is updated recursively for $n = 1, \ldots, N$ and the final value is obtained as κ_N. By default, the center of the hyperellipsoid is used to construct the stegosignal (equation 3.3) and the embedded watermark vector is $\omega = \mathbf{a}_c(N) - \mathbf{a}$.

The watermark recovery process for SMF parametric watermarking involves LSE estimation of the modified LP coefficients (refer Table 3.2). Hence, even in case of SMF-based watermarking, the embedded watermarks are asymptotically ($N \rightarrow \infty$) immune to an additive white noise attack [47].

4.3 Robustness optimization

The robustness property is dependent on selection of appropriate watermark solution from the hyperellipsoidal set, strength of the embedded watermark, and watermark detection. In general, greater robustness can be obtained by embedding more energetic watermarks and this in turn, affects stegosignal fidelity. Although, by default, the center of the hyperellipsoid constitutes the watermark solution, in most cases it might not be the optimal solution for a given attack. For robustness analysis it is assumed that the hyperellipsoidal set $\mathcal{E}_N$ is obtained through the SMF filtering algorithm subjected to the fidelity constraint, $|\tilde{y}_n - y_n| < \gamma_n$. It should be noted that the hyperellipsoid is not centered at $\mathbf{a}$, the vector of original LP coefficients. More energetic watermark vectors are embedded by selecting perturbed LP parameters from $\mathcal{E}_N$ that are as further away as possible from the original LP parameters $\mathbf{a}$. The selection of appropriate watermark solution from $\mathcal{E}_N$ depends on the attack and the targeted robustness.

The SMF-based watermarking approach is especially useful in improving watermark robustness against attacks whose effects vary based on the nature of the watermark signal. For example, robustness to a lowpass filtering attack can be improved by selecting low frequency watermark signals.

4.3.1 Optimal watermarks for a filtering attack

The impulse response of an attack filter is assumed to be known and is denoted by $\{h_n\}$. The stegosignal of form (equation 3.3) is to be constructed by selecting an appropriate vector of perturbed LP coefficients from the hyperellipsoidal set $\boldsymbol{\mathcal{E}}_N$. The corresponding watermark vector is defined as $\omega = \tilde{\mathbf{a}} - \mathbf{a}$. Let $\{\tilde{y}_n^f\}_{n=1}^N$ be the result of filtering the stegosignal. That is, at time n,

$$\tilde{y}_n^f = \tilde{y}_n * h_n = y_n * h_n + w_n * h_n, \tag{4.7}$$

where $\{\tilde{y}_n\}$ is assumed to be a stegosignal constructed from any $\tilde{\mathbf{a}} \in \boldsymbol{\mathcal{E}}_N$ including the best (optimized) $\bar{\mathbf{a}}$. An ineffective attack on the stegosignal will result in a filtered stegosignal with a filtered coversignal component that is perceptually dissimilar to the original. This is because watermark robustness is generally defined as the ability of the watermark to survive an attack to the extent that the speech fidelity is not affected beyond an application-dependent criterion. Also, an attack is ineffective if the filtered watermark signal $\{w_n^f\}_{n=1}^n$ approximates the original watermark signal $\{w_n\}_{n=1}^n$. The coversignal and the attack filter $\{h_n\}_{n=1}^n$ are predetermined quantities and hence the filtered coversignal component in equation (4.7) cannot be controlled by the watermark embedding algorithm. However, the second term in (4.7) ($\{w_n^f\}_{i=1}^n$) can

be made to be robust against the filtering attack by selecting an appropriate $\bar{\mathbf{a}}$ from the set $\boldsymbol{\mathcal{E}}_N$. The problem of selecting the "best" set of modified LP coefficients from $\boldsymbol{\mathcal{E}}_N$, is now addressed.

Let Δw_n^f be defined as,

$$\begin{aligned}\Delta w_n^f &= w_n^f - w_n \\ &= h_n * w_n - w_n \\ &= \sum_i [(\tilde{a}_i - a_i) y_{n-i}] * (h_n - \delta_n),\end{aligned}$$

where δ_n is the Kronecker delta; $\delta_0 = 1$ and $\delta_n = 0$ for $n \neq 0$. Clearly, Δw_n^f is a function of $\tilde{\mathbf{a}}$ for a given attack filter. Then the mean squared error (MSE) between the filtered and original watermark signals is given by,

$$f(\tilde{\mathbf{a}}) = E(\Delta w_n^f)^2 = \frac{1}{N} \sum_{n=1}^{N} (w_n^f - w_n)^2. \tag{4.8}$$

If $\bar{\omega} = \bar{\mathbf{a}} - \mathbf{a}$ is indeed the "best" watermark vector, then the corresponding filtered watermark signal $\bar{\mathbf{w}}^f$ is associated with minimum MSE. Then, $\bar{\mathbf{a}}$ is obtained by solving the following constrained optimization problem:

$$\begin{aligned} &\text{minimize} \quad f(\tilde{\mathbf{a}}) \\ &\text{subject to} \quad \tilde{\mathbf{a}} \in \boldsymbol{\mathcal{E}}_N \end{aligned} \tag{4.9}$$

The method of lagrange multipliers can be used to solve this optimization problem [28]. The domain of the constraint function is the

hyperellipsoid, which is a convex set if $\frac{\mathbf{C}_N}{\kappa_N}$ is a positive definite matrix.

4.3.2 Optimal watermarks for a quantization attack

This section deals with uniform and non-uniform scalar quantizer attacks on watermarks. The quantizer consists of L equal or unequal intervals $[I_1, I_2, \ldots, I_L]$. Each interval I_l, for $l = 1, 2, \ldots, L$ is associated with a quantization value x_l. The scalar quantization operation Q can be expressed as,

$$Q(\tilde{y}_n) = \tilde{y}_n^q$$

where $\tilde{y}_n^q = x_l$ whenever $|\tilde{y}_n^q - x_l|$ is minimum over $l = 1, 2, \ldots, L$.

To maximize watermark robustness to a specific quantization attack, a similar constrained optimization problem to that in equation (4.9) is solved with the objective function $f(\tilde{\mathbf{a}}) = \frac{1}{N}\sum(w_n^q - w_n)^2$, where $w_n^q = \tilde{y}_n^q - y_n$. In a similar way, optimal watermarks for best robustness to a combination of filtering and quantization attacks can be determined. The latter problem can be generalized for a combined attack involving several distinct attacks on the stegosignal.

4.3.3 Maximizing watermark energy

The boundary of the hyperellipsoidal set obtained by SMF considerations is given by,

$$\boldsymbol{\mathcal{E}}_N^b = \{\tilde{\mathbf{a}} | (\tilde{\mathbf{a}} - \mathbf{a}_c(N))^T \frac{\mathbf{C}_N}{\kappa_N} (\tilde{\mathbf{a}} - \mathbf{a}_c(N)) = 1\}, \tag{4.10}$$

where $\mathbf{a}_c(N)$ is the center of the hyperellipsoid. The boundary of the hyperellipsoid is significant for the following reason. An important factor affecting watermark robustness is the energy of the watermark signal ($w_n = \tilde{y}_n - y_n$). The modified LP coefficients from the boundary of the hyperellipsoid result in the highest energy watermarks for the corresponding fidelity constraint. Watermark robustness is also a function of the frequency content of the watermark signal. However, this paper is mainly concerned with the effect of watermark signal energy on watermark robustness and the "best" watermark vector is selected accordingly. The "best" watermark vector $\bar{\omega}$ is such that the corresponding vector of modified LP coefficients $\bar{\mathbf{a}}$ is from the hyperellipsoidal boundary $\boldsymbol{\mathcal{E}}_N^b$. The constrained optimization problem in (4.9) is modified as follows.

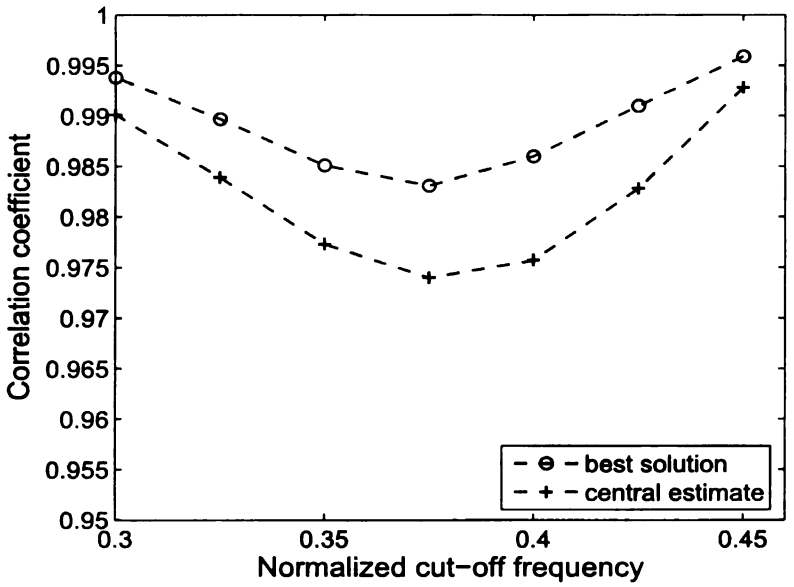
$$\begin{aligned} &\text{minimize} \quad f(\tilde{\mathbf{a}}) \\ &\text{subject to} \quad \tilde{\mathbf{a}} \in \boldsymbol{\mathcal{E}}_N^b \end{aligned} \tag{4.11}$$

4.4 Experiments and discussion

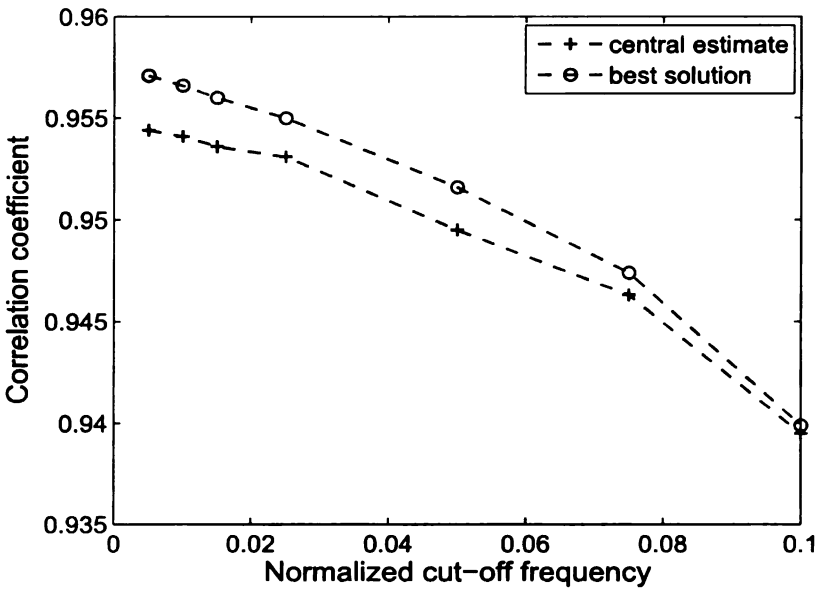
Although Lagrange multipliers can be used to solve the optimization problem in (4.9), the subsequent computational complexity might be too costly for certain watermarking applications. Moreover, this research is mainly concerned with selecting modified LP coefficients such that the resulting watermark signal has high energy. Hence, searching the hyperellipsoidal boundary at intermittent points for improved robustness will prove to be beneficial. There is a trade-off between the number of points selected from the hyperellipsoidal set and computational complexity. As an example, the experiments reported in this section were executed in Matlab running on a 1.4GHz Celeron processor with 512 MB RAM and an average run time of 5 seconds.

Experiments were performed to test the robustness of the "best" SMF solution to filtering and quantization attacks. The coversignal consisted of 500 samples of the vowel sound /A/ sampled at 10 kHz. The correlation coefficient between the original and recovered watermarks is used as a measure of robustness. A 4^{th}order LP model was used for watermarking. The value of γ_n was 0.4 for all $n \in [1, N]$, and the watermark signal was imperceptible in the resulting stegosignal.

Figure 4.1(a) shows the effect of a low pass filtering attack involving a 4^{th}order lowpass Butterworth on the best SMF solution and the cen-



(a)



(b)

Figure 4.1: Filtering attack. For Fig. 4.1(a) a 4^{th}order IIR Butterworth lowpass filter was used to distort the stegosignal, and for Fig. 4.1(b) an 8^{th}order FIR highpass filter was used to attack the stegosignal.

tral estimate (default solution) of the membership set. In Fig. 4.1(b), a highpass filtering attack was applied on the stegosignal by using an 8^{th}order FIR filter. The watermarks derived from the best solution (ellipsoid boundary) are more robust than those derived from the central estimate of the set. It is seen from Table 4.1 that parametric watermarking is quite robust to quantization attacks. The original coversignal was quantized at 16 bits per sample. The uniform quantizer in the attack used 3 bits per sample. A sub-optimal non-uniform quantizer requiring 3 bits to code the quantized value was implemented by arbitrarily partitioning the quantization range. Finally, Fig. 4.2 shows the effect of both quantization and lowpass filtering on recovered watermarks derived from the best and default solutions of the hyperellipsoidal set. In almost all cases the watermarks recovered from the best SMF solution perform significantly better than the ones recovered from modified LP coefficients at the center of the membership set.

In applications with little prior knowledge of potential attacks, diversity [43] in watermark embedding is employed for improved robustness. The SMF robustness optimization can be viewed in this context for embedding multiple watermarks, each with a targeted robustness to specific combination of attacks.

Table 4.1: Robustness to quantization attacks

Type of quantizer	**SMF solution**	**central estimate**
	corr coef	corr coef
uniform	1	0.9998
Non-uniform	0.9998	0.9990

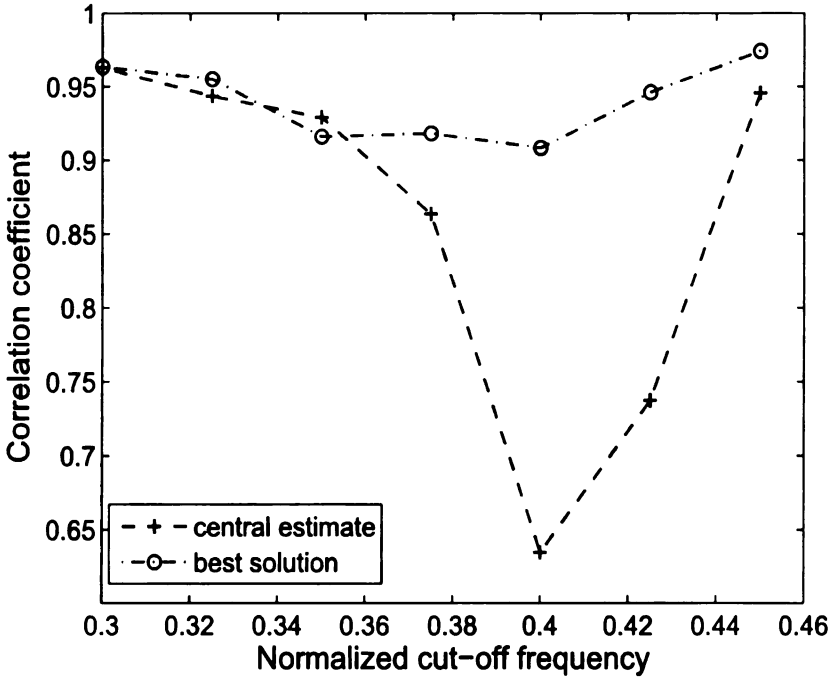


Figure 4.2: Watermark robustness to combination of non-uniform quantization and IIR lowpass filtering attacks.

Chapter 5

Generalizations and Extensions

5.1 Introduction

In this chapter, a generalized framework for speech watermarking based on linear-in-parametric models of speech production process is presented. Watermarks are embedded in the LSP parameters, log area ratio (LAR) parameters, inverse sine (IS) parameters, and reflection coefficients (parcor coefficients) [6]. The watermark robustness and stegosignal fidelity aspects of these alternate parametric speech models are discussed in this chapter and compared with watermarking in the LP domain.

The chapter also presents an application of perturbed parameter theory to watermarking [48, 49]. The perturbed parameter theory is used for obtaining bounds on the perturbation of the stegosignal caused by watermarking, hence in assessing the effects of the embedded water-

marks on fidelity.

5.2 Generalized framework for parametric watermarking

A consequence of the LP watermarking framework is that alternate or related representations of LP parametric models can be used for watermarking. These representations, including LAR, LSP, IS, and parcor coefficients, may prove to be beneficial for watermarking in certain applications. For example, localization of watermark content in the frequency domain is more effectively controlled through direct manipulation of LSP coefficients. On the other hand, since LAR coefficients have the highest correlation with subjective quality [50], they can be directly altered to preserve stegosignal fidelity.

In order to obtain the LSP parameters, the Z-domain representation of the M^{th} order LP inverse filter is decomposed into the following polynomials.

$$P(Z) = A(Z) + Z^{-(M+1)}A(Z^{-1}) \tag{5.1}$$

$$Q(Z) = A(Z) - Z^{-(M+1)}A(Z^{-1}) \tag{5.2}$$

$$A(Z) = \frac{P(Z) + Q(Z)}{2} \tag{5.3}$$

where $A(Z) = 1 - \sum_{i=1}^{M} a_i Z^{-i}$, the Z-domain representation of the in-

Table 5.1: Generalized watermark embedding algorithm

Let $\{y_n\}_{n=-\infty}^{\infty}$ denote a coversignal, and let $\{y_n\}_{n=n_k}^{n_k'}$ be the k^{th} of K speech frames to be watermarked. Then: For $k = 1, 2, \ldots, K$

1 Using the "autocorrelation method" (e.g., [6, Ch. 5]), derive a set of LP coefficients of order M, say $\{a_i\}_{i=1}^{M}$, for the given frame.

2 Use the LP parameters in an *inverse filter* configuration to obtain the prediction residual on the frame, $\left\{\xi_n = y_n - \sum_{i=1}^{M} a_i y_{n-i}\right\}_{n=n_k}^{n_k'}$.

3 Convert the LP parameters to LSP or parcor parameters and embed the watermark vectors. Alternately, for watermarking in the LAR or IS domain, convert the LP parameters to parcor and then convert the resulting parcor parameters into LAR or IS parameters before embedding the watermark vectors. Use the modified LSP, LAR, IS, or parcor parameters to produce a corresponding set of modified LP parameters, say $\{\tilde{a}_i\}_{i=1}^{M}$.

4 Use the modified LP parameters as a (suboptimal) predictor of the original sequence, adding the residual obtained in Step 2 above at each n, to resynthesize the speech over the frame, $\left\{\tilde{y}_n = \sum_{i=1}^{M} \tilde{a}_i y_{n-i} + \xi_n\right\}_{n=n_k}^{n_k'}$. (To the extent that the watermark represents only small perturbations to the original LP parameters, the resynthesized result is a pointwise approximation to the coversignal over the same time frame.)

5 The sequence $\{\tilde{y}_n\}_{n_k}^{n_k'}$ is the k^{th} frame of the watermarked speech (stegosignal).

Next k.

verse filter [6]. The zeros of the polynomials P and Q constitute the LSP parameters. The zeros of P and Q occur in complex conjugate pairs and hence M unique zeros are required to specify the vocal tract model [6]. The magnitude of the zeros is unity and only the frequency parameter is required to be represented. The LSPs represent the frequency parameters. Conversion from LP domain to LSP or LSP to LP parameters is quite simple [equations (5.1), (5.2) and (5.3)]. The watermark embedding and recovery algorithms presented in Tables 5.1 and 5.2 include LSP to LP and LP to LSP conversions respectively, for watermarking in the LSP domain. Since LSPs represent frequencies of zeros lying within the unit circle, it has to be ensured that the modified LSP parameter values are within 0 and π. This requirement imposes a constraint on the strength of the embedded watermark vectors and consequently on the energy of the watermark signal.

The reflection coefficients (κ) constitute an alternate representation to LP coefficients and play an important role in speech coding and analysis applications. The parcor coefficients are obtained as a by-product of the Levinson-Durbin (L-D) recursion, which is used to convert the autocorrelation values of speech to LP coefficients. Conversion from reflection coefficients to LP coefficients is accomplished using the algorithm in Table 5.3 [51]. LP coefficients can be converted to reflection coefficients using the algorithm in Table 5.4 [51]. Water-

Table 5.2: Generalized watermark recovery algorithm

For $k = 1, 2, \ldots, K$

1 Subtract residual frame $\{\xi_n\}_{n_k}^{n'_k}$ from the stegosignal frame $\{\tilde{y}_n\}_{n_k}^{n'_k}$. This results in an estimate of the modified predicted speech, $\{d_n = \tilde{y}_n - \xi_n\}_{n_k}^{n'_k}$.

2 Estimate the *modified* LP coefficients $\{\tilde{a}_i\}_1^M$ by computing the least-square-error solution, say $\{\hat{\tilde{a}}_i\}_1^M$, to the overdetermined system of equations: $d_n \approx \sum_{i=1}^{M} \alpha_i y_{n-i}$, $n = n_k, \ldots, n'_k$.

3 Convert the modified LP coefficients from Step 2 to modified LSP, LAR, IS, or parcor coefficients.

4 Use the parameter estimates from Step 3 to derive the corresponding watermark values.

Next k.

mark information can be added to the reflection coefficients and the resultant converted to modified LP coefficients. While embedding the watermark, it should be ensured that $|k_i| \neq 1$ for any i, otherwise finding the reflection coefficients is an ill-conditioned problem.

Other sets of speech parametric models for embedding watermark information include the LAR and inverse sine parameters. The LAR and inverse sine parameters are related to the reflection coefficients as shown in equations (5.4) and (5.5), respectively:

$$\upsilon_l = \frac{1}{2} \log \frac{1 + \kappa_l}{1 - \kappa_l} = \tanh^{-1} \kappa_l, \text{ for } l = 1, 2, \cdots, M. \qquad (5.4)$$

Table 5.3: Conversion of reflection coefficients to LP coefficients

Let κ be a vector of M reflection coefficients.

1 Initialize the output LP vector $\mathbf{a}$ to the first element of κ, i.e., κ_1.

2 For $i = 2, \cdots, M$, $\mathbf{a} = \mathbf{a} + [(a_{i-1}, \cdots, a_1) * \kappa_i, \kappa_i]$. Next i.

3 The final set of LP coefficients are obtained as the final vector $\mathbf{a}$

Table 5.4: Conversion of LP coefficients to reflection coefficients

For $j = M, \cdots, 1$,

1 Let $\kappa_j = a_j$.

2 Consider elements 1 through j of $\mathbf{a}$. Let $\mathbf{a} = [a_1, \cdots, a_j]$ and let $\underline{\mathbf{a}} = [a_j, a_{j-1}, \cdots, a_1]$.

$$\mathbf{a} = (\mathbf{a} - \kappa_j \underline{\mathbf{a}})/(1 - \kappa_j^2)$$

Next j.

$$\Upsilon_l = \frac{2}{\pi} \sin^{-1} \kappa_l, \text{ for } l = 1, 2, \cdots, M. \tag{5.5}$$

Information can be embedded by modifying these parameters. The modified LAR or inverse sine parameters are converted to the corresponding modified reflection coefficients, which in turn are converted to modified LP coefficients. The stegosignal is reconstructed by following steps 4 and 5 in Table 5.1.

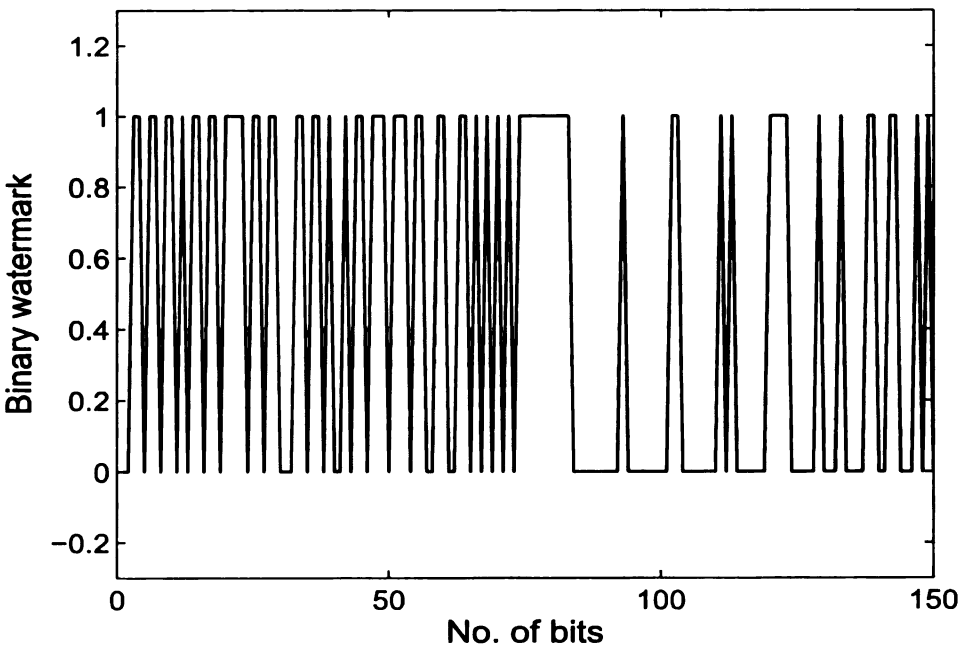


Figure 5.1: The first 100 bits of the 1000-bit binary watermark.

5.3 Experiments and discussion

Experiments were performed to compare the robustness and fidelity aspects of watermarking in the LSP, LAR, IS, and parcor domains with LP watermarking. Speech was watermarked in the LP domain using the algorithm in Table 3.1. The algorithm in Table 5.1 was used for watermarking in the LSP, LAR, IS, and parcor domains.

In the experiments presented below, the coversignal consists of 15.625 s of speech from the TIMIT database [41], sampled at 16 kHz. The coversignal consisted of samples from ten different sentences of the TIMIT database, uttered by a female talker. The first 24000 or 26000

samples of the ten sentences were watermarked. A 1000-bit pseudo random binary watermark sequence was generated [Fig. 5.1]. The coversignal was divided into 125 frames of 2000 samples or 0.125 seconds duration each. In each of the speech frames, a length eight watermark vector was embedded into the coefficients of an eighth-order parametric model, resulting in a data payload of 64 bits per second. Very few audio watermarking algorithms can satisfactorily trade-off robustness and fidelity at a payload of 43 bits per second. For the LP parametric watermarking experiments presented in this section, the watermark embedding and recovery do not involve selective normalization. The sample correlation coefficient [equation 3.21] is used as the measure of similarity between original and recovered watermark vectors for all the parametric watermarking techniques.

5.3.1 Subjective perceptual tests

Parametric watermarking in LP, LAR, IS, and parcor domains was implemented at CWR_{seg}'s of 7.715 dB, 10.68 dB, 27 dB, and 30 dB. LSP-based parametric watermarking was implemented at CWR_{seg}'s of 27 dB and 30 dB. In LSP-based watermarking, the modified LSP parameters must be between 0 and π and this imposes a constraint on watermarking at higher CWR_{seg}'s of 7.715 dB and 10.68 dB. Parameter-embedded watermarks were fairly inaudible at these CWR_{seg} [30].

Although CWR_{seg} is used as the objective measure of fidelity, listening tests were also performed to compare the watermarked speech fidelity. Five subjects were asked to rank the stegosignals from different parameter-embedded watermarking schemes in terms of the perceptual similarity to the coversignal. The sound files used in the subjective listening tests are available at the website [30]. At CWR_{seg} of 7 dB, the stegosignal from LP watermarking was found to have the highest fidelity followed by stegosignals from LAR and IS watermarking. At CWR_{seg} of 7 dB, the stegosignal from parcor watermarking was found to have the least fidelity. At CWR_{seg} of 27 dB, the fidelity of LP, LSP, LAR, and IS stegosignals was comparable. While, the stegosignal fidelity of parcor watermarking was marginally worse.

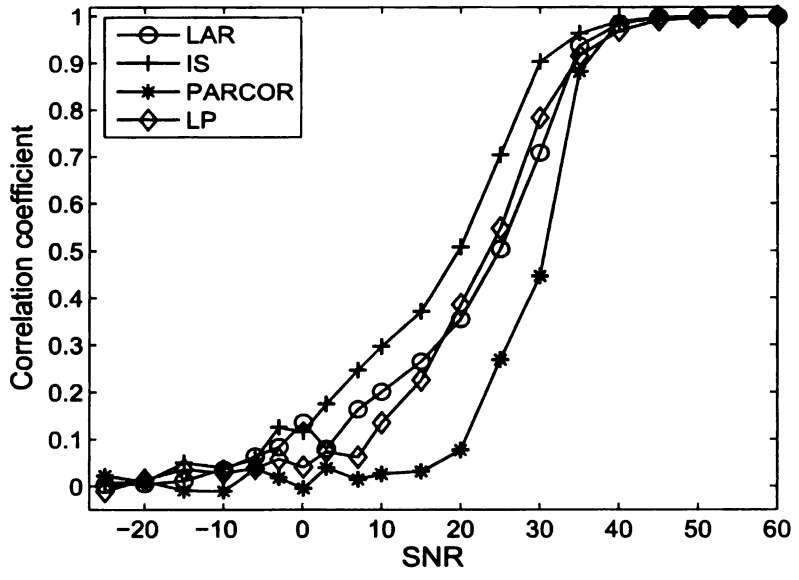
5.3.2 Robustness experiments

Experiments were performed to study the robustness of LP, LSP, LAR, IS, and parcor based watermarking algorithms to additive noise, and speech coding. The stegosignals were subjected to white Gaussian noise in the time domain resulting in SNRs ranging from -30 dB to 60 dB. Figure 5.2(a) shows the correlation coefficient between the 1000-bit original and recovered watermarks for LP, LAR, IS, and parcor watermarking at CWR_{seg} of 7 dB. It can be observed from Fig. 5.2(a) that parcor watermarking is least robust to additive noise. And IS water-

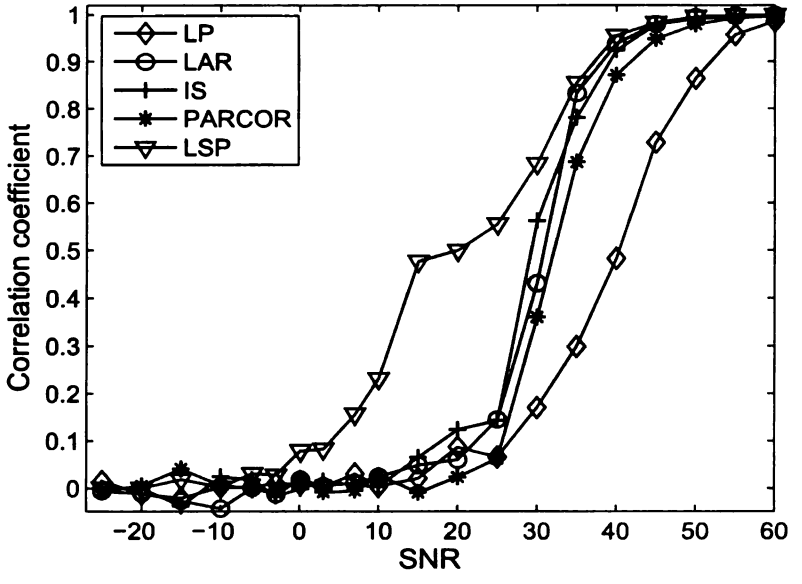
marking results in better robustness than LP and LAR watermarking at CWR_{scg} of 7 dB. The improved robustness of LP watermarking at 7 dB CWR_{seg} in Fig. 3.5 compared to LP watermarking in Fig. 5.2(a) is due to selective normalization. The robustness of LP, LSP, LAR, IS, and parcor watermarking to additive white noise at CWR_{seg} of 27 dB is shown in Fig. 5.2(b). At 27 dB CWR_{scg}, LSP-based watermarking results in more robust watermarks and LP watermarking without normalization results in the least robust watermarks.

Robustness of LP, LSP, LAR, IS, and parcor watermaking to existing speech coding schemes was tested. G.711, ADPCM, GSM, CELP and LPC10 were the speech coders used for robustness testing [22]. The original coversignal and stegosignal consisted of 256k bits per second. The stegosignals were coded (compressed), then decoded (decompressed), and the watermarks were recovered from the decoded signals. The correlation coefficient between the original and recovered watermarks for different speech coders are tabulated in Tables 5.5 and 5.6 for the different parametric watermarking techniques. The attacked stegosignals are available in the website [30]. The coding bit rates of the different speech coders are also listed in Tables 5.5 and 5.6.

It is seen from Tables 5.5 and 5.6, that all the attacked stegosignals at CWR_{seg} of 7 dB and 27 dB are highly robust to the G.711 μ-law coder, except LP watermarking at 27 dB CWR_{scg}. LSP watermarking



(a)



(b)

Figure 5.2: Effect of white Gaussian noise on LP, LSP, LAR, IS and PARCOR embedded watermarks. In 5.2(a) a CWR_{seg} of 7 dB was used to obtain the stegosignals, and in 5.2(b) a CWR_{seg} of 27 dB was used to obtain the stegosignals.

Table 5.5: Robustness to speech coding CWR_{seg} of 7 dB

Speech codec	**Bit rate**	**LP 7 dB**	**LAR 7 dB**	**IS 7 dB**	**PARCOR 7 dB**
	k bits/s	corr coef	corr coef	corr coef	corr coef
G.711	64	0.9916	0.9844	0.9967	0.9962
ADPCM	32	0.6669	0.8502	0.9381	0.6746
GSM	13.2	0.0246	0.1151	0.1481	−0.0074
CELP	4.5	−0.0905	−0.0180	−0.0043	−0.0250
CELP	2.3	−0.0665	0.0043	0.0291	−0.0078
LPC10	2.4	−0.0825	−0.0266	0.0194	0.0281

is highly robust to G.711 μ-law and G.726 ADPCM coders even at a higher CWR_{seg} of 27 dB. The robustness of all the parametric schemes decreases for very low bit rate speech coding [13.2k to 2.3k bits per second]. A similar trend was observed in the robustness of SS watermarking to low bit rate speech coding in Table 3.5.

Watermark robustness to very low bit rate speech coding can be improved by compromising watermark payload for greater robustness. For example, error correcting stategies can be applied to watermarking [52] and the watermark can be repeatedly embedded for greater diversity [43].

Table 5.6: Robustness to speech coding at CWR_{seg} of 27 dB

Speech codec	Bit rate	LP 27 dB	LAR 27 dB	IS 27 dB	PARCOR 27 dB	LSP 27 dB
	k bps	corr coef	corr coef	corr coef	corr coef	corr coef
G.711	64	0.7767	0.9868	0.9840	0.9662	0.9966
ADPCM	32	0.0578	0.7590	0.7247	0.6212	0.8663
GSM	13.2	−0.0662	0.0225	0.0146	0.0034	0.0872
CELP	4.5	−0.0898	0.0068	0.0125	−0.0310	0.0124
CELP	2.3	−0.0639	0.0065	0.0453	−0.0357	−0.0207
LPC10	2.4	−0.0927	0.0049	−0.0176	−0.0319	−0.0217

5.4 Perturbed parameter models in watermarking

Deller and Gulboy [48, 49], determined conditions under which an autoregressive (AR) model[1] with stochastic parameters can be approximated by a time-invariant one wherein the stochastic coefficients are replaced by their mean values. In this section, the application of AR perturbed parameter theory to LP parametric speech watermarking is explored. We consider a general Markov equation with slightly perturbed parameters of the form

$$\tilde{\mathbf{y}}_{n+1} = \Phi(\tilde{\delta}_n, \tilde{\mathbf{y}}_n, n). \tag{5.6}$$

In general, $\tilde{\mathbf{y}}_n$ is Q-vector, $\tilde{\delta}_n$ is a first order stationary stochastic R-vector, and Φ is a general vector function of $\tilde{\delta}$, $\tilde{\mathbf{y}}$ and n. The conditions

[1]The AR model is the statistician's name for an LP model driven by white noise

under which the model in equation (5.6) is well approximated by the following model [equation (5.7)] are given in [48,49]:

$$\mathbf{y}_{n+1} = \phi(\mathbf{y}_n, n), \text{ where } \phi(\tilde{\mathbf{y}}_n, n) = E[\Phi(\tilde{\delta}_n, \tilde{\mathbf{y}}_n, n)|\tilde{\mathbf{y}}_n, n]. \tag{5.7}$$

It is shown that, if the following conditions are true with probability one on $n \in [1, N]$:

$$\|\Phi(\tilde{\delta}_n, \tilde{\mathbf{y}}_n, n) - \phi(\tilde{\mathbf{y}}_n, n)\| \leq \epsilon, \tag{5.8}$$

$$\|\phi(\tilde{\mathbf{y}}'_n, n) - \phi(\tilde{\mathbf{y}}''_n, n)\| \leq K_L \|\tilde{\mathbf{y}}'_n - \tilde{\mathbf{y}}''_n\|, \tag{5.9}$$

then the models in equations (5.6) and (5.7) approximate according to,

$$\|\tilde{\mathbf{y}}_n - \mathbf{y}_n\| \leq \epsilon \left\{ 1 + \sum_{i=1}^{N-1} K_L^i \right\}, \text{ w.p.1, } n \in [1, N]. \tag{5.10}$$

This perturbed parameter theory is used for obtaining bounds on the perturbation of the coversignal caused by watermarking.

5.4.1 Time-varying AR models in watermarking

The stegosignal obtained using equation (3.3) is manipulated into a time-varying AR model:

$$\begin{aligned}\tilde{y}_n &= \sum_{i=1}^{M} \tilde{a}_i y_{n-i} + \xi_n \\ &= \sum_{i=1}^{M} \left[a_i + \omega_i + \frac{(a_i + \omega_i)(y_{n-i} - \tilde{y}_{n-i})}{\tilde{y}_{n-i}} \right] \tilde{y}_{n-i} + \xi_n \\ &= \sum_{i=1}^{M} \tilde{\tilde{a}}_{n,i} \tilde{y}_{n-i} + \xi_n.\end{aligned}$$

The expression for the time-varying AR coefficients $\{\tilde{\tilde{a}}_{n,i}\}$ can be manipulated into the form

$$\tilde{\tilde{a}}_{n,i} = a_i \rho_{n,i} + \omega_i \rho_{n,i} = a_i + a_i(\rho_{n,i} - 1) + \omega_i \rho_{n,i},$$

where $\rho_{n,i} = \left(\frac{y_{n-i}}{\tilde{y}_{n-i}}\right) \approx 1$. The time-varying AR parameters are composed of the true parameter term, a_i, and the perturbation term, $a_i(\rho_{n,i} - 1) + \omega_i \rho_{n,i}$.

5.4.2 Application of perturbed parameter Markov equations to watermarking

Now let us suppose that the stegosignal is constructed such that,

$$\tilde{y}_n = \sum_{i=1}^{M} (a_i + \omega_{n,i}) y_{n-i} + \xi_n = \sum_{i=1}^{M} \tilde{a}_{n,i} y_{n-i} + \xi_n, \tag{5.11}$$

where $\omega_{n,i} = \begin{cases} \omega_i, & \text{if } n \text{ is odd} \\ -\omega_i, & \text{if } n \text{ is even} \end{cases}$. Then the time-varying AR parameters are given by,

$$\tilde{\tilde{a}}_{n,i} = \tilde{a}_{n,i}\rho_{n,i} = (a_i + \omega_{n,i})\rho_{n,i}. \tag{5.12}$$

The AR(M) system is written in state space formulation as follows. Let,

$$\left.\begin{aligned} \tilde{\mathbf{y}}_{n+1} = \tilde{\mathbf{A}}_n\tilde{\mathbf{y}}_n + \mathbf{G}\xi_n = \Phi(\tilde{\delta}_n, \tilde{\mathbf{y}}_n, n) \\ \tilde{y}_n = \mathbf{c}^T\tilde{\mathbf{y}}_{n+1} \end{aligned}\right\}. \tag{5.13}$$

where

$$\tilde{\mathbf{A}}_n = \left|\begin{array}{c|c} \tilde{\tilde{a}}_{n,1} \quad \tilde{\tilde{a}}_{n,2} \quad \cdots & \tilde{\tilde{a}}_{n,M} \\ \hline \mathbf{I}_{(M-1)\times(M-1)} & \begin{matrix} 0 \\ \vdots \\ 0 \end{matrix} \end{array}\right|,$$

$$\tilde{\mathbf{y}}_n = [\tilde{y}_{n-1}, \tilde{y}_{n-2}, \cdots, \tilde{y}_{n-M}]^T,$$

$$\mathbf{G} = \mathbf{c} = [1, 0, \cdots, 0]^T,\ \tilde{\delta}_n = [\tilde{\tilde{a}}_{n,1}, \tilde{\tilde{a}}_{n,2}, \cdots, \tilde{\tilde{a}}_{n,M}]^T.$$

The watermark coefficients ($\{\omega_{n,i}\}_{i=1}^{M}$) are such that the time-varying AR coefficients are first order stationary with $E[\tilde{\tilde{a}}_{n,i}] = a_i$ and $|\tilde{\tilde{a}}_{n,i} - a_i| < \iota$. The perturbed AR parameter theory is used for determining

how well the AR model in equation (5.13) is approximated by the model

$$\left.\begin{aligned} \mathbf{y}_{n+1} = \mathbf{A}_n\mathbf{y}_n + \mathbf{G}\xi_n = \phi(\mathbf{y}_n, n) \\ y_n = \mathbf{c}^T\mathbf{y}_{n+1} \end{aligned}\right\}. \tag{5.14}$$

In the above equation, $\mathbf{A} = E[\tilde{\mathbf{A}}_n]$. The vector $\mathbf{y}_n$ is defined similarly to the analogous vector in equation (5.13). As demonstrated in [48], it is similarly determined that the small perturbation condition ($|\tilde{\tilde{a}}_{n,i} - a_i| < \iota$) is equivalent to the condition (5.8) of the theorem.

$$\|\Phi(\tilde{\delta}_n, \tilde{\mathbf{y}}_n, n) - \phi(\tilde{\mathbf{y}}_n, n)\|_* = \|(\tilde{\mathbf{A}}_n - \mathbf{A})\tilde{\mathbf{y}}_n\|_* \leq \|\tilde{\mathbf{A}}_n - \mathbf{A}\|_*\|\tilde{\mathbf{y}}_n\|_* \tag{5.15}$$

According to [48, 49], the matrix norm $\|.\|_*$ is selected such that for any square matrix $\mathbf{A}$, $\|\mathbf{A}\|_* \leq r(\mathbf{A}) + e$, given any $e > 0$. Here, $r(\mathbf{A})$ represents the spectral radius of the matrix $\mathbf{A}$. In Lemma 5.6.10 of [53] the matrix norm $\|.\|_*$ is given by

$$\|\mathbf{A}\|_* = \|\mathbf{D}_t\mathbf{U}^T\mathbf{A}\mathbf{U}\mathbf{D}_t^{-1}\|_1 = \|(\mathbf{U}\mathbf{D}_t^{-1})^{-1}\mathbf{A}(\mathbf{U}\mathbf{D}_t^{-1})\|_1, \tag{5.16}$$

in which $\|.\|_1$ is the maximum column sum matrix norm induced by the ℓ_1 vector norm, and $\mathbf{D}_t = \text{diag}(t, t^2, t^3, \ldots, t^n)$ with $t > 0$ and sufficiently large. The matrix $\mathbf{U}$ is obtained by the Schur decomposition of $\mathbf{A}$ given by $\mathbf{A} = \mathbf{U}\Delta\mathbf{U}^T$, Δ being an upper triangular matrix with the main diagonal components comprised of the eigenvalues of

$\mathbf{A}$. The vector norm compatible with the induced matrix norm $\|.\|_*$ is the ℓ_1 norm since $\|\mathbf{D}_t\mathbf{U}^T\mathbf{A}\mathbf{U}\mathbf{D}_t^{-1}\mathbf{x}\|_1 \leq \|\mathbf{D}_t\mathbf{U}^T\mathbf{A}\mathbf{U}\mathbf{D}_t^{-1}\|_1\|\mathbf{x}\|_1$ and $\mathbf{D}_t\mathbf{U}^T\mathbf{I}\mathbf{U}\mathbf{D}_t^{-1} = \mathbf{I}$, the identity matrix.

It is reasonable to assume that $\|\xi_n\|_*$ is bounded by W, a non-negative finite number. The bound on $\|\tilde{\mathbf{y}}_n\|_*$ is determined in [48, 49] and is given by,

$$\|\tilde{\mathbf{y}}_n\|_* \leq Y_0 \prod_{k=0}^{n-1} \|\tilde{\mathbf{A}}_k\|_* + \sum_{k=0}^{n-2} \left[\prod_{j=k+1}^{n-1} \|\tilde{\mathbf{A}}_j\|_* \right] \|\mathbf{G}\|_* W + \|\mathbf{G}\|_* W \quad (5.17)$$

Where, $Y_0 = \|\tilde{\mathbf{y}}_0\|_*$ and $\|\mathbf{G}\|_* = 1$. Also, $\tilde{\mathbf{A}}_n \approx \mathbf{A}$ for all n and hence it is assumed that there exists a small number $e'(\iota)$, a function of ι, such that $\|\tilde{\mathbf{A}}_n\|_* \leq r(\tilde{\mathbf{A}}_n) + e'(\iota)$. Let p be the maximum pole magnitude of the system associated with $\mathbf{A}$ and hence $p = r(\mathbf{A})$. Similarly, let $p(n)$ be the pointwise maximum pole magnitude associated with $\tilde{\mathbf{A}}_n$ and $p(n) = r(\tilde{\mathbf{A}}_n)$ [48, 49]. Also, $p(n) < p + \alpha\iota$ for small ι and α being a constant. Then, $\|\tilde{\mathbf{A}}_n\|_* \leq p + \alpha\iota + e'(\iota)$. Since $|\tilde{\tilde{a}}_{n,i} - a_i| < \iota$, $\|\tilde{\mathbf{A}}_n - \mathbf{A}\|_* \leq f(\iota)$ for some number $f(\iota)$. Hence, equation (5.17) can be rewritten as,

$$\|\tilde{\mathbf{y}}_n\|_* \leq Y_0(p + \alpha\iota + e'(\iota))^n + W\left\{1 + \sum_{l=1}^{n-1}(p + \alpha\iota + e'(\iota))^l\right\} = S(n). \quad (5.18)$$

Also, $S(n) < \infty$ and $S(n) < S(N)$ for all $n \in [1, N]$. Hence equation

(5.15) can be expressed as,

$$\|\Phi(\tilde{\delta}_n, \tilde{\mathbf{y}}_n, n) - \phi(\tilde{\mathbf{y}}_n, n)\|_* \leq f(\iota)S(N) = \epsilon(N), \qquad (5.19)$$

with probability one. The above equation represents condition (5.8) of the theorem. The existence of Lipschitz constant K_L in equation (5.9) can be demonstrated in a similar way as in [48,49],

$$\phi(\tilde{\mathbf{y}}_n, n) - \phi(\mathbf{y}_n, n) = (\mathbf{A}\tilde{\mathbf{y}}_n + \mathbf{G}\xi_n) - (\mathbf{A}\mathbf{y}_n + \mathbf{G}\xi_n) = \mathbf{A}(\tilde{\mathbf{y}}_n - \mathbf{y}_n).$$

Hence, condition (5.9) of the theorem is given by,

$$\|\phi(\tilde{\mathbf{y}}_n, n) - \phi(\mathbf{y}_n, n)\|_* \leq \|\mathbf{A}\|_* \|\tilde{\mathbf{y}}_n - \mathbf{y}_n\|_* = (p+e)\|\tilde{\mathbf{y}}_n - \mathbf{y}_n\|_*, \quad (5.20)$$

with probability one. Based on conditions (5.19) and (5.20), the norm of the difference between the stegosignal and the coversignal is bounded as follows.

$$\|\tilde{\mathbf{y}}_n - \mathbf{y}_n\|_* \leq (f(\iota)S(N)) \left\{ 1 + \sum_{i=1}^{N-1} (p+e)^i \right\}, \qquad (5.21)$$

with probability one. Experiments were performed on speech data to determine the final bounds on the watermark signal for parameter perturbations. The coversignal consisted of 20 mS of the vowel sound /A/ sampled at 10000 Hz. An 8^{th}order LP inverse filter was used for

watermarking and the watermark vector was,

$$\omega = 10^{-3} * [-0.1025, -0.1234, 0.0289, -0.0429, 0.0056, -0.0368, -0.0465, 0.0371].$$

By applying the perturbed parameter theory, the right-hand side of equation (5.21) was determined to be 11.03. It was verified experimentally that the ℓ_1 norm of the difference between the stegosignal and coversignal is bounded by 11.03. A tighter upper bound would be of greater significance to practical implementations of parametric watermarking. A relatively high value of 11.03 is obtained in the right-hand side of (5.21) as the parameter perturbations are of higher energy than the underlying requirement in the formulation of the theorem [conditions (5.8) and (5.9)].

Chapter 6

Conclusions

The dissertation presents a general approach to watermarking of speech signals based on LP, LSP, LAR, IS, and PARCOR parametric models. The dissertation focusses on embedding watermark information by directly or indirectly modifying the long-term LP parameters of speech. Parametric watermarking incorporates characteristics of SS watermarking algorithms, as well as those of integration-by-synthesis techniques. These aspects strongly influence the fidelity, security and robustness characteristics of the technique.

Watermark recovery is treated as a system identification problem involving LSE estimation. The watermark information is concentrated during the embedding and recovery phases, while it is temporally and spectrally distributed otherwise. The distributed nature of the watermark combined with the LSE estimation during recovery, contribute to watermark robustness.

The dissertation initially focussed on speech watermarking in the LP domain. In experiments presented here, and in many others, LP parametric watermarking has proven to be robust to most common forms of attack. An example parametric watermark detector has been presented to assess performance. The noise in the parameter domain was found to be Gaussian distributed when white or colored noise was added to the stegosignal in the time domain. By selectively normalizing watermark coefficients to parameter magnitudes, $1/|a_i|$, whenever $|a_i| > 1$, the parameter noise affecting the watermark coefficients was rendered independent of the original predictor coefficients. Through this selective normalization, watermark detection can be treated as signal detection problem in the presence of Gaussian noise. Very low false-alarm rates are achieved.

The method of Lagrange multipliers can be used for obtaining optimally robust watermarks from perturbed LP coefficients selected from the membership set. SMF optimization is however not useful against attacks that are independent of the stegosignal. For applications limited by computational complexity and where the energy of the watermark signal is considered to be of main significance to robustness, searching the hyperellipsoidal boundary at intermittent points results in more robust watermarks than the central estimate of the membership set. The use of SMF in obtaining robust watermarks to filtering, quantization,

and combination attacks is demonstrated.

The fidelity and robustness aspects of LP, LSP, LAR, IS, and parcor parametric watermarking algorithms were compared. It is determined that stegosignals obtained by LP and LAR watermarking are generally associated with high fidelity even at a low CWR_{seg} of 7 dB. Although LSPs cannot be watermarked at 7 dB CWR_{seg}, LSP-based watermarking is highly robust to noise and G.711 and G.726 codecs even at a CWR_{seg} of 27 dB. In general, parametric watermarking is much less robust to CELP and LPC10 codecs compared to G.711, G.726 and GSM codecs. However, the quality of speech decompressed by low bit rate CELP and LPC10 codecs is very low.

An application of AR perturbed parameter theory to speech watermarking is presented and bounds are obtained on the watermark signal for small parameter perturbations.

Parametric watermarking algorithms can be used for applications such as content management, broadcast monitoring, owner identification and copyright protection. Parametric watermarking is highly robust to additive noise, quantization errors, speech codecs such as G.711, G.726, GSM, and cropping. Based the requirements of an application, the fidelity and robustness of parameter-embedded watermarks can be systematically adjusted.

Bibliography

[1] M.S. SEADLE, J.R. DELLER, JR. and A. GURIJALA, "Why watermark? The copyright need for an engineering solution," *Proceedings of ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, Portland, July 2002.

[2] I.J. COX, M.L. MILLER and J.A. BLOOM, *Digital Watermarking*, Academic Press, 2002.

[3] N.F. JOHNSON, Z. DURIC and S. JAJODIA, *Information Hiding: Steganography and Watermarking - Attacks and Countermeasures*, Kluwer Academic Publishers, 2000.

[4] S. VOLOSHYNOVSKIY, S. PEREIRA, T. PUN, J.K. SU and J.J. EGGERS, "Attacks and benchmarking," *IEEE Communications Magazine*, August 2001.

[5] A. GURIJALA and J.R. DELLER, JR., "Robust algorithm for watermark recovery from cropped speech," *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, May 2001.

[6] J.R. DELLER, JR., J.H.L. HANSEN and J.G. PROAKIS, *Discrete-Time Processing of Speech Signals* (2d ed.), IEEE Press, 2000.

[7] A. GURIJALA, J.R. DELLER, JR., M.S. SEADLE and J.H.L. HANSEN, "Speech watermarking through parametric modeling," *Proceedings of International Conference on Spoken Language Processing (ICSLP)*, Denver, CO, September 2002.

[8] S. HAYKIN, *Adaptive Filter Theory* (3d ed.), Prentice-Hall, 1996.

[9] J.H.L. HANSEN, B. ZHOU, M. AKBACAK, R. SARIKAYA and B.L. PELLOM, "Audio stream phrase recognition for a National Gallery of the Spoken Word: One small step," *Proceedings of ICSLP*, Beijing, October 2000, pp. 1089-1092.

[10] I.J. COX, J. KILIAN, T. LEIGHTON and T. SHAMOON, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1673-1687, December 1997.

[11] F.J. RUIZ and J.R. DELLER, JR., "Digital watermarking of speech signals for the national gallery of the spoken word," *Proceedings of IEEE ICASSP*, Istanbul, Turkey, May 2000, pp. 1089-1092.

[12] D. ANAND and U.C. NIRANJAN, "Watermarking medical images with patient information," *Proceedings of IEEE/EMBS Conference*, Hong Kong, October 1998, pp. 703-706.

[13] S.G. MIAOU, C.H. HSU, Y.S. TSAI and H.M. CHAO, "A secure data hiding technique with heterogeneous data-combining capability for electronic patient records," *Proceedings of the World Congress on Medical Physics and Biomedical Engineering: Electronic Healthcare Records*, Chicago, July 2000.

[14] T. KALKER, G. DEPOVERE, J. HAITSMA and M. MAES, "A video watermarking system for broadcast monitoring," *Proceedings of SPIE IS&T/SPIE's* 11th *Annual Symposium on Electronic Imaging '99: Security and Watermarking of Multimedia Contents*, Chicago, January 1999, vol. 3657.

[15] A.S. SPANIAS, "Speech Coding: A Tutorial Review," *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1541-1582, October 1994.

[16] Q. CHENG and J. SORENSEN, "Spread spectrum signalling for speech watermarking," *Proceedings of IEEE ICASSP*, Salt Lake City, May 2001, vol. 3, pp. 1337-1340.

[17] M. HAGMÜLLER, H. HORST, A. KRÖPFL and G. KUBIN, "Speech watermarking for air traffic control," *Proceedings of* 12th*European Signal Processing Conference*, Vienna, Austria, September 2004.

[18] M. HATADA, T. SAKAI, N. KOMATSU and Y. YAMAZAKI, "Digital watermarking based on process of speech production," *Proceedings of SPIE: Multimedia Systems and Application*, 2002, vol. 4861.

[19] M. CELIK, G. SHARMA and A.M. TEKALP, "Pitch and Duration Modification for Speech Watermarking," *Proceedings of IEEE ICASSP*, Philadelphia, PA, March, 2005, vol. 2, pp. 17-20.

[20] E. MOLINES and F. CHARPENTIER, "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," *Speech Communication*, vol. 9, no. 5-6, pp. 453-467, December 1990.

[21] B. CHEN and G.W. WORNELL, "Quantization index modulation: A class of provably good methods for digital watermarking and information embedding," *IEEE Transactions on Information Theory*, vol. 47, no. 4, pp. 1423-1443, May 2001.

[22] A.M. KONDOZ, *Digital Speech: Coding for Low Bit Rate Communication Systems* (2d ed.), John Wiley & Sons, 2004.

[23] S. GOLLAMUDI, S. NAGARAJ, S. KAPOOR and Y.F. HUANG, "SMART: A toolbox for set-membership filtering," *Proceedings of 1997 European Conference on Circuit Theory and Design*, Budapest, Hungary, 1997.

[24] S. NAGARAJ, S. GOLLAMUDI, S. KAPOOR and Y.F. HUANG, "BEACON: An adaptive set-membership filtering technique with sparse updates," *IEEE Transactions on Signal Processing*, vol. 47, no. 11, pp. 2928-2941, November 1999.

[25] J.R. DELLER, JR. and Y.F. HUANG, "Set-membership identification and filtering for signal processing applications," *Circuits, Systems, and Signal Processing. (Special issue on signal processing and its applications)*, vol. 21, no. 1, pp. 69-82, January 2002.

[26] J.R. DELLER, JR., M. NAYERI and S.F. ODEH, "Least square identification with error bounds for real-time signal processing and control," *Proceedings of the IEEE*, vol. 81, pp. 813-849, June 1993.

[27] J.R. DELLER, JR., "Set membership identification in digital signal processing," *IEEE Acoustics, Speech and Signal Processing Magazine*, vol. 6, no. 4, pp. 4-20, October 1989.

[28] S. BOYD and L. VANDENBERGHE, "Convex Optimization," *Cambridge University Press*, 2004.

[29] A. GURIJALA and J.R. DELLER, JR., "Speech Watermarking by Parametric Embedding with an ℓ_∞ Fidelity Criterion," *Proceedings of Eurospeech-2003*, Geneva, Switzerland, September 2003, pp. 2933-2936.

[30] SPEECH FILES, http://www.egr.msu.edu/~deller/ParaWmkg/WAVFILES.

[31] D. GRUHL, A. LU and W. BENDER, "Echo hiding," *Lecture Notes in Computer Science; Proceedings of the First International Workshop on Information Hiding*, Cambridge, UK, 1996, vol. 1174, pp. 293-315.

[32] S. WANG, A. SEKEY and A. GERSHO, "An objective measure for predicting subjective quality of speech coders," *IEEE Journal on Seclected Areas in Communications*, vol. 10, no. 5, pp. 819-829, June 1992.

[33] S.A. CRAVER, N. MEMON, B-L. YEO and M. YEUNG, "Resolving Rightful Ownerships with Invisible Watermarking Techniques: Limitations, Attacks, and Implications," *IEEE Journal of Selected Areas in Communications - Special issue on Copyright and Privacy Protection*, vol. 16, no. 4, pp. 573-586, May 1998.

[34] J.J. HERNANDEZ, M. AMADO and F. PEREZ-GONZALEZ, "DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure," *IEEE Transactions on Image Processing*, vol. 9, no. 1, pp. 55-68, January 2000.

[35] M. BARNI, F. BARTOLINI, A.D. ROSA and A. PIVA, "Optimal decoding and detection of multiplicative watermarks," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1118-1123, April 2003.

[36] T.P. CHEN and T. CHEN, "A framework for optimal blind watermark detection," *Proceedings of ACM Multimedia and Security Workshop*, Ottawa, Canada, October 2001.

[37] J.P.M.G. LINNARTZ, A.C.C. KALKER and G.F. DEPOVERE, "Modeling the false-alarm and missed detection rate for electronic watermarks," *Lecture Notes in Computer Science*, vol. 1525, pp. 329-343, Springer-Verlag, 1998.

[38] M.L. MILLER and J.A. BLOOM, "Computing the probability of false watermark detection," *Proceedings of the Third Workshop on Information Hiding*, Dresden, Germany, 1999, pp. 146-158.

[39] A. GURIJALA and J.R. DELLER, JR., "Detector design for parametric speech watermarking," *IEEE International Conference on Multimedia and Expo (ICME)*, Amsterdam, The Netherlands, July 2005, pp. 251-255.

[40] H.V. POOR, *An Introduction to Signal Detection and Estimation* (2d ed.), Springer-Verlag, 1994.

[41] P.J. PRICE, "A database for continuous speech recognition in a 1000-word domain," *Proceedings of IEEE ICASSP*, New York, vol. 11, pp.651-654, 1988.

[42] G.R. COOPER and C.D. MCGILLEM *Modern Communications and Spread Spectrum*, McGraw-Hill Book Company, 1996.

[43] D. KUNDUR and D. HATZINAKOS, "Diversity and attack characterization for improved robust watermarking," *IEEE Transactions on Signal Processing*, vol. 29, no. 10, pp. 2383-2396, October 2001.

[44] J.G. PROAKIS and D.G. MANOLAKIS, *Digital Signal Processing: Principles, Algorithms, and Applications* (3rd ed.), Prentice-Hall, 1996.

[45] F.A.P. PETITCOLAS, R.J. ANDERSON and M.G. KUHN, "Attacks on copyright marking systems," *Proceedings of Second Workshop on Information Hiding*, Portand, Oregon, April 1998, pp.218-238.

[46] HAWKVOICE FROM HAWK SOFTWARE, http://www.hawksoft.com/hawkvoice.

[47] A. GURIJALA and J.R. DELLER, JR., "Speech watermarking with objective fidelity and robustness criteria," *Proceedings of Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2003.

[48] J.R. DELLER, JR. and Z. GULBOY, "Simplified models for perturbed parameter Markov equations with application to ARMA systems," *International Journal on Systems Science*, vol. 14, no. 10, pp. 1185-1190, 1983.

[49] J.R. DELLER, JR. and Z. GULBOY, "A correction to 'Simplified models for perturbed parameter Markov equations with application to ARMA systems'," *International Journal on Systems Science*, vol. 15, no. 8, pp. 915-916, 1984.

[50] S.R. QUACKENBUSH, T.P. BARNWELL and M.A. CLEMENTS, *Objective Measures of Speech Quality*, Prentice-Hall, NJ, 1988.

[51] S. KAY, *Modern Spectral Estimation: Theory and Application*, Prentice-Hall signal processing series, NJ, 1988.

[52] S. BAUDRY, J.-F. DELAIGLE, B. SANKUR, B. MACQ and H. MAITRE "Analysis of error correction strategies for typical communication channels in watermarking," *Signal Processing*, vol. 81, pp. 1239-250, 2001.

[53] R.A. HORN and C.R. JOHNSON, *Matrix Analysis*, Cambridge University Press, 1996.