This is to certify that the

thesis entitled

ROBUSTNESS OF TEC SPEECH WATERMARKING TO CROPPING AND ADDITIVE NOISE

presented by

Aparna Gurijala

has been accepted towards fulfillment
of the requirements for

Master's degree in Electrical Eng

John R. Deller Jr
Major professor

Date 04 May 2001

ROBUSTNESS OF TEC SPEECH WATERMARKING TO CROPPING AND ADDITIVE NOISE

By

Aparna Gurijala

A THESIS

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

MASTER OF SCIENCE

Department of Electrical and Computer Engineering

2001

ABSTRACT

ROBUSTNESS OF TEC SPEECH WATERMARKING TO CROPPING AND ADDITIVE NOISE

By

Aparna Gurijala

The widespread use of the internet has created a need for technologies for the protection of copyrighted digital information. Digital watermarking is one such technology in which a preferably imperceptible signal (*watermark*) is embedded into a copyrighted host signal. Digital watermarks are prone to a wide range of "attacks" and other forms of distortion. In this work, the robustness of a new watermarking method based on transform encryption coding (TEC) to cropping and additive noise is investigated.

Experiments were conducted to test the robustness of TEC speech watermarking to additive noise under different conditions including different SNRs and watermark masking algorithm parameters. Although a cropping attack is easy to implement, the resulting desynchronization severely hinders watermark detection and recovery. A dynamic programming (DP) based algorithm for the detection of cropped speech samples and reconstruction of the cropped stego-signal to enable watermark recovery has been developed. Implementation details of the DP algorithm and performance under different environmental conditions are presented. Factors influencing the robustness of TEC speech watermarking are analyzed.

ACKNOWLEDGMENTS

I would like to acknowledge Dr. J.R. Deller, my advisor for his invaluable guidance, encouragement and support. Special thanks to Dr.Deller for his very helpful remarks and suggestions that greatly contributed to my learning and understanding. Special thanks to Dr.Seadle and Dr.Radha for their consideration, patience and effort. The time spent by Dr.Deller, Dr. Seadle and Dr.Radha to ensure the completion of my thesis is truly appreciated.

Personally I would like to thank my parents for their love and encouragement. My thanks to all my friends for their kindness and help.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Watermarking for the National Gallery of the Spoken word

The National Gallery of the Spoken Word (NGSW) [1] project is creating an online database of spoken word collections, spanning the 20th century. These collections are mainly drawn from Michigan State University's Vincent Voice Library, MSU Museum, Chicago Historical Society and Northwestern University. They include Thomas Edison's first cylinder recordings to the voices of Theodore Roosevelt, Florence Nightingale, and Babe Ruth. The aural resources for the NGSW are in the digital form.

Representation of information in digital form has many properties that make it preferable to analog forms. An unlimited number of digital copies can be made with ease and accuracy. This benefit, however, has been a cause of concern for intellectual property owners and content providers. The widespread use of the Internet coupled with the developments in compression techniques facilitates fast and efficient distribution of digital content. However, while easy to implement, distribution of copyrighted digital information without authorization threatens intellectual property rights. Copyright laws protecting analog information are inapplicable to digital information. As a result, there is a need to develop techniques for protecting the ownership of digital content and for tracking intellectual piracy.

Digital watermarking is one such technique. Digital watermarking is the process of embedding a permanent and preferably imperceptible signal into a copyrighted host signal. The embedded signal may typically convey information about the owner, author

or carrier. More information about the need for watermarking in the NGSW project is found in [7].

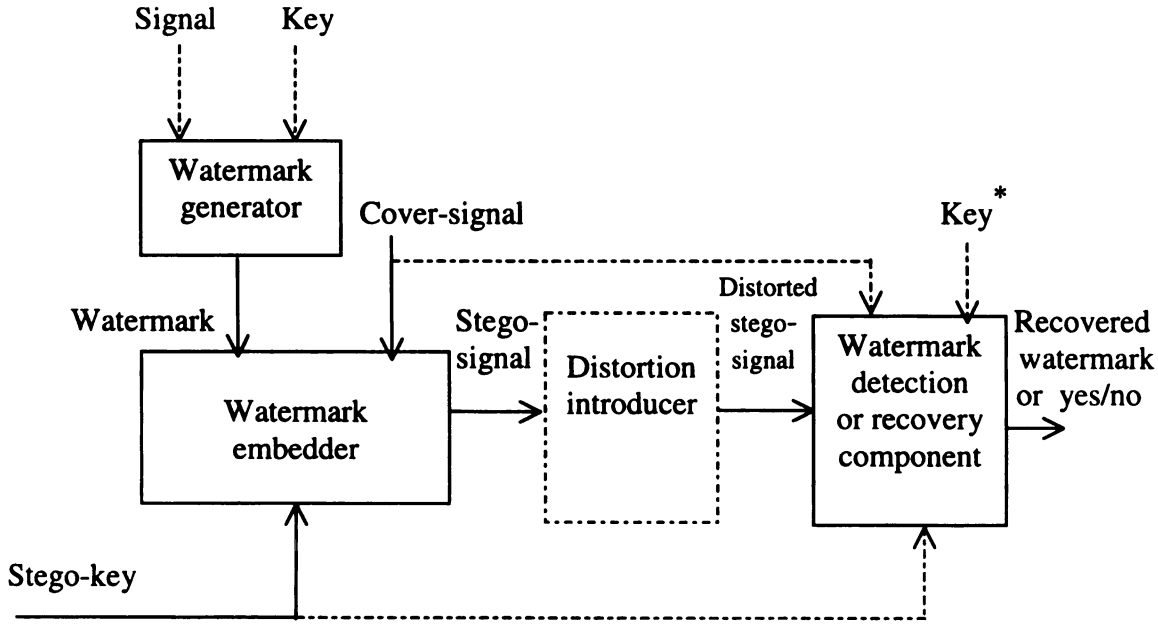
The concept of watermarking has its origins in the ancient Greek technique of steganography or "covered writing" – interpreted as hiding information in other information. Detailed information on the history of steganography and watermarking is found in [8]. Applications of digital watermarking include copyright protection, fingerprinting, authentication, copy control, owner identification, broadcast monitoring, security control and tamper proofing. Watermarking can be used to protect virtually any form of digital information including images, speech, music, and video.

Most of the digital watermarking schemes have been developed for images. Audio watermarking schemes include the method due to Boney *et al.* [9] in which the watermark is generated by filtering a PN-sequence with a filter that approximates the frequency masking characteristics of the human auditory system, and then accounting for temporal masking. Bassia and Pita [10] developed an audio watermarking method that modifies the temporal characteristics of the audio signal in accordance with a *seed* (watermark *key*) known only to the copyright owner. In [11] an audio watermarking technique operating in the Fourier domain is presented. Bender *et al.* [12] use homomorphic signal processing techniques to place information imperceptibly into audio streams by the introduction of closely spaced echoes. Luy *et al.* [13] proposed a multi-purpose audio watermarking scheme that embeds two complementary watermarks – one for audio authentication and the other for the detection of tampered regions. The spread spectrum watermarking technique developed by Cox *et al.* [14] can be applied to audio, image, video and multimedia data.

This paper is concerned with the robustness of the digital speech watermarking technique employing *transform encryption coding* (TEC) [2, 3].

1.2 A typical watermarking system

A typical watermarking system consists of a watermark generator, an embedder, a watermark detector and possibly a component that distorts the stego-signal (defined below).



* Same as the one used for watermark generation.

Figure 1. A typical watermarking system

Due to the wide variations in the watermarking techniques, it is difficult to generalize and characterize a "typical" watermarking scheme. To account for the vast variations in watermarking approaches, certain inputs are indicated by dotted lines in Figure 1, meaning that they may not be present in all techniques. A signal for which copyright protection must be provided is called a *cover-signal.* A *watermark* is a signal

that is embedded into the cover-signal for this purpose in accordance with the *stego-key*[1]. The stego-key ensures the imperceptibility of the watermark and thus introduces additional protection, by making the watermark location unknown. A watermark may take different forms – an encrypted or modulated speech sequence or image, least significant bit manipulations, a pseudo-random sequence. As a result, the inputs to watermark generators are highly diverse. For example, in the audio watermarking technique proposed by Bassia and Pitas [10] the input signal is the cover-signal itself, and the key is a randomly generated constant. In the spread spectrum watermarking scheme of Cox *et al.* [14], the input signal is the same as the key and comprises a pseudo-random sequence.

Watermark embedding techniques may be additive, multiplicative or quantization-based [15] and may operate in the space or time domain, or in some transform domain. The output of a watermark embedder is the *stego-signal.* The stego-signal should be perceptibly similar to the cover-signal, in spite of the presence of the watermark. *Watermark detectors* are classified as type I or type II. Type I detectors require knowledge of the cover-signal to extract the watermark from the stego-signal. Type II detectors provide a yes or no answer to question of whether the watermark is present in a distorted stego-signal. In the motivating application for this work, the TEC speech watermarking system employs a type I detector.

Typically the term "watermark" is used to refer to the processed (modulated, encrypted, etc.) form of the original signal to be embedded in the cover-signal as indicated in Figure 1. However, in this document, "watermark" will refer to the

[1] In the TEC speech watermarking technique the stego-key is the constant or adaptive gain factor of the masking algorithm [2].

unprocessed watermark signal. The result of the processing step will be called the *encrypted watermark.*

1.3 Properties of digital watermarks

Some essential properties of watermarks are as follows:

- *Perceptual transparency:* Inserting a watermark into the host or cover-signal will alter the cover-signal in some way. If the amount of alteration does not introduce any perceptual degradation then the watermark is said to be *perceptually transparent* [16,17]. This ensures that the value of the original material is not reduced by the presence of the watermark.
- *Robustness:* Robustness refers to the degree to which a watermark can survive an "attack" or distortion. An *attack* is a deliberate attempt to remove the watermark or hinder its recovery. The watermark should not be able to be destroyed without simultaneous destruction of the cover-signal. A successful attack is one that removes the watermark or obstructs the recovery process without causing perceptual degradation of the cover-signal [16].
- *Unambiguity:* A recovered watermark should unambiguously identify the owner of the watermarked material.
- *Security:* Encryption keys, if any, used in the watermarking process, and keys used for watermark generation, should be very difficult to predict, guess, or otherwise ascertain.

Another important property is the watermark bit rate [17]. This is determined by the amount of information contained in the watermark (watermark *payload*) and the

amount of data needed to embed one unit of watermark information (watermark *granularity*) while ensuring perceptual transparency.

For greater robustness it is desirable to have stronger components of the watermark in the stego-signal. This in turn will affect the perceptual transparency of the signal containing the watermark (stego-signal). Thus there are trade-offs among the various watermark properties that must be considered in light of the requirements of a particular watermarking application. Further, for an application like fragile watermarking [19] robustness is not desirable. In such a case, fragile watermarks that get destroyed by some or all of the transformations are used. The degree of adherence to the ideal properties is dictated by the requirements of the particular application and the availability of resources. More information is found in [16]-[19].

1.4 Classification of watermarking techniques

Watermarking techniques are classified according to the domain in which the watermark is inserted, the requirements of the watermark detection process, or the availability of the keys.

Watermarking schemes are categorized as *restricted-* or *unrestricted-key* watermarking schemes based on the relative availability of the key(s) [20]. Schemes in which the keys are available to all the watermark detectors are called unrestricted-key schemes. In the case of restricted-key schemes, the knowledge of keys is confined to a small number of detectors. The TEC-based speech watermarking scheme is a restricted key scheme. Though such a categorization appears to be mainly based on a difference in usage, the complexity and suitability of a watermarking algorithm differs between the two cases.

Schemes that require knowledge of the cover-signal to recover the watermark are said to be *non-oblivious* [21]-[23]. TEC-based watermarking is non-oblivious. Watermark recovery is effected by subtracting the cover-signal from the stego-signal. Non-oblivious techniques generally yield more robust watermarks. However, non-oblivious watermarking may be more prone to *protocol attacks* [20]-[22] due to the availability of greater freedom for creating fake cover-signals and hence fake watermarks. For example, a hacker may succeed in developing a suitable fake watermark (say, a pseudo-random pattern). On subtracting it from the stego-signal, to which he or she has access, a fake original can be created. The hacker now claims to be the owner of this original. Of course, oblivious watermarking schemes are more prone to attacks based on neutralizing the detector (if there is access to one as in the case of the DVD copy control problem [20]) readings. The cover-signal is not required during the detection process in *oblivious* watermarking and may be treated as noise. Oblivious watermarking methods permit faster detection of the watermark and include bit-wise or noise-dependent methods. These methods are sensitive to even small variations of the stego-signal and are thus more fragile [22]

A watermarking strategy is designated as a spatial (time) or transform domain technique according to whether the watermark is embedded into the cover-signal in the signal or the transform domain. In the present application in which audio rather than image data are watermarked, the term "signal," rather than "spatial," domain is more appropriate.

If the same key is required in the watermark recovery or detection process as that used for watermark embedding, the scheme is said to be *symmetric*. The need for

asymmetric or *public key* watermarking arises when the *user* of the copyrighted information [23] must perform watermark detection. In this case there is a set of two keys – a public key and a private key. The private key is required for watermark embedding and recovery, and is known only to the owner. The public key is given to the users solely for watermark *detection*. Knowledge of the public key should not provide any information about the private key, and should not compromise the security and robustness of the scheme. A variation on this idea occurs in the TEC strategy. A "public" key is made available to descramble the speech signal, but this process has the effect of further encrypting the watermark rather than detecting it.

1.5 Attacks on watermarking systems

Digital watermarks are prone to a wide range of attacks [15, 20] and other means of distortion. As mentioned earlier, an attack is an attempt to remove the watermark or preclude its recovery, while ensuring tolerable or no apparent damage to the stego-signal. An attack can also be an attempt to create ambiguity of ownership. Attacks include those due to common signal processing operations like resampling, compression, filtering, D/A conversion, and requantization. Introduction of noise can also affect a watermark. Deliberate manipulations of the content like cropping, rescaling and rotation can severely hinder watermark recovery.

By using secure keys, a cryptographic attack like *brute-force key search* [25] can be thwarted. In the attack by *statistical averaging* [18, 20, 25], a large number of differently watermarked copies of the same stego-signal may be averaged to get the attacked stego-signal. *Collusion attack* differs from an attack by statistical averaging in

the sense that only portions of the stego-signal and not the entire stego-signal are used to create the attacked stego-signal [25, 26].

Counterfeit attacks [15, 24, 25], including inversion attack, multiple watermarks, and copy attack, attempt to undermine the concept of watermarking itself by producing fake originals or fake watermarked signals. Watermarking an already marked signal (the problem of multiple watermarks) can negate the utility of any watermarking scheme. To counteract these attacks, watermark registration with a trusted authority has been proposed by some parties [21, 23].

Distortion is normally the result of signal processing operations or the presence of noise. Attacks encompass the different types of distortion that may be unintentionally introduced into the stego-signal.

Robustness against attacks is a very important aspect of a watermarking scheme. A particular watermarking scheme may not be robust to all forms of attack. An attack on the watermark may be directed at either removing the watermark or hindering its recovery while causing tolerable apparent damage to the stego-signal. When the attack hinders watermark recovery, then the general remedy is to attempt to identify the attack and to undo the damage. Duric *et al.* [22] describe a method of recognizing distorted images and recovering watermarks using identification marks, salient features of the image invariant to transformations like cropping, scaling and rotation. In the present paper, watermark recovery from cropped speech is accomplished using a dynamic programming approach. Most watermarking techniques are susceptible to damage caused by cropping due to its desynchronization of the watermark detection and recovery process.

1.6 Document overview

Research in the field of watermarking is progressing in different directions. New watermarking techniques are being devised [11, 13, 27], new attacks on watermarking schemes are being identified [15, 18, 20, 25], benchmarks to evaluate the different watermarking schemes are being developed [15, 28], and algorithms for watermark detection and recovery after attacks and other forms of distortion are being developed [22, 24]. This document describes work that was mainly directed at TEC-based watermark detection and recovery after being subjected to certain attacks.

Chapter 2 presents the TEC-based speech-watermarking technique. Chapter 3 describes the robustness of this watermarking scheme to different attacks that include additive noise, cropping, or a combination. Algorithms for watermark recovery when subjected to these attacks are described. Chapter 4 is concerned with Matlab implementation, experimental results and performance evaluation under different conditions. Chapter 5 comprises a description of future work.

Chapter 2

DIGITAL WATERMARKING OF SPEECH USING TEC

2.1 Watermarking algorithm

Transform encryption coding (TEC) was originally developed by Kuo *et al.* [3] as an algorithm for image compression and efficient and secure transmission. It can be applied to speech and other signals as well. TEC produces independent transform coefficients by passing the signal through an all-pass filter with unity gain. TEC derives its encryption properties, and hence security, from the use of highly random filter coefficients. Typically, quasi m-arrays and gold code arrays [4] are used to obtain filter coefficients with the desired property of unpredictability. The phase spectrum of the signal to be transformed is scrambled in accordance with the phase spectrum of a quasi m-array or gold code array [3].

The speech watermarking technique developed by Ruiz *et al.* [2] employs TEC in conjunction with a masking algorithm for encrypting and watermarking speech. The one dimensional speech signal is arranged in the form of two-dimensional arrays, each having the same dimensions as the quasi m-arrays used for the TEC process.

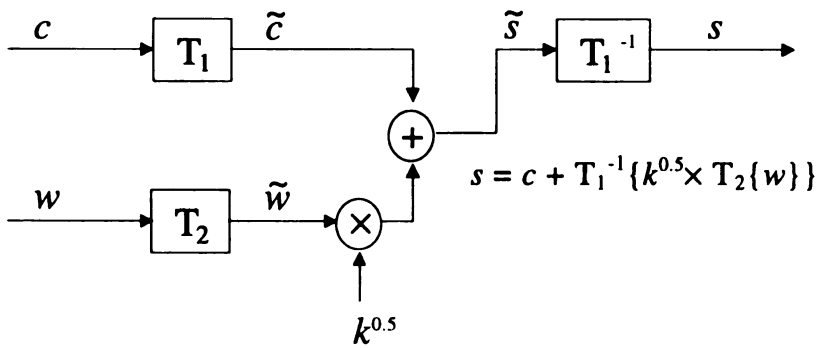


Figure 2. Watermarking process

The watermarking process involves the application of TEC to both the cover-signal and the watermark. Different quasi m-arrays are used for encrypting the cover-signal and the watermark. The encrypted watermark is subjected to a masking algorithm to ensure perceptual transparency based on the cover to watermark ratio (CWR), defined as

$$CWR_{dB} = 10\log_{10}\frac{\tilde{C}[n]}{k[n]\times\tilde{W}[n]} \tag{1}$$

where $\tilde{C}[n]$ and $\tilde{W}[n]$ are the respective short-term energy measures for the encrypted cover and watermark signals, and $k[n]$ is an adaptive gain factor (*stego-key*) at time n. Alternately, a constant gain factor k can be used instead of $k[n]$. Since the encryption process involves passing the cover and watermark signals through all-pass filters with unity gain [see Figure 4], the energy content of the encrypted and non-encrypted signals are similar in each case. The encrypted cover-signal and watermark are converted into a one-dimensional arrays. The encrypted, masked watermark is then added to the encrypted cover-signal to obtain the encrypted *stego-signal*. Applying the inverse TEC operation to decrypt the cover-signal component of the encrypted stego-signal subjects the watermark to a second level of encryption (see Figure 6).

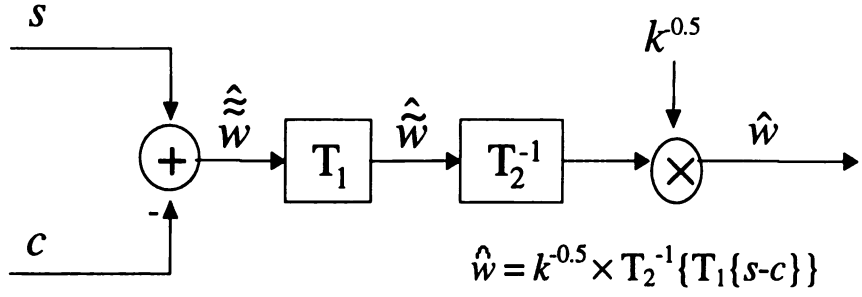


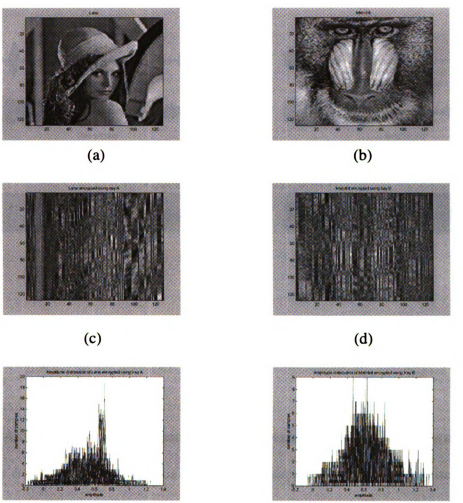
Figure 3. Watermark recovery

For watermark recovery, an estimate of the doubly encrypted watermark is obtained by subtracting the cover-signal from the stego-signal,

$$\hat{\tilde{w}} = s - c \tag{2}$$

Finally the inverse TEC operations and the gain factor are applied to the estimated twice-encrypted watermark (Figure 3):

$$\hat{w} = k^{-1} \times \mathrm{T}_2^{-1}\left\{\mathrm{T}_1\left\{\hat{\tilde{w}}\right\}\right\} = k^{-1} \times \mathrm{T}_2^{-1}\left\{\mathrm{T}_1\left\{s - c\right\}\right\} \tag{3}$$

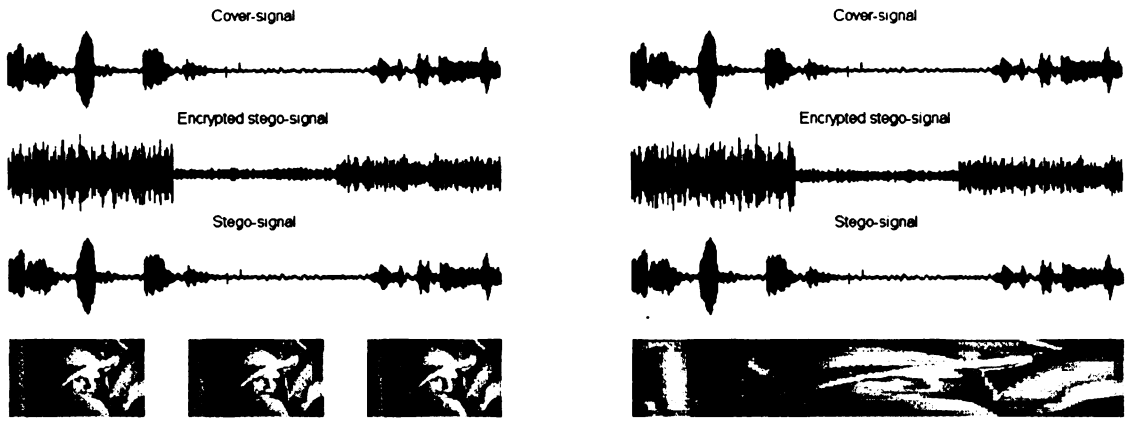


(a) (b) (c) (d) (e) (f)

Figure 4. Encryption using quasi m-arrays. (a) The original "Lena" image. (b) The original "mandrill" image. (c) Lena encrypted using quasi m-array A (key A). (d) Mandrill encrypted using quasi m-array B (key B). (e) Amplitude distribution of the Lena

encrypted using key A. (f) Amplitude distribution of the mandrill encrypted using key B. The amplitude distributions are different for the encrypted versions of the two images. However, they are similar to the amplitude distributions of the respective original images due to the all pass nature of the encryption process.

The recovery of the watermark is only possible with the knowledge of the two quasi m-arrays (encryption keys) used in the process. The watermark may take many forms including speech samples or images. Part of the future work of this project will be concerned with researching properties that assure quality watermarks.



(a) (b)

Figure 5. Watermarking selectively to watermarking the entire speech record. (a) The entire speech record consisting of 48387 samples in watermarked. All the recovered watermarks are shown. (b) The watermark is embedded in the first 16129 speech samples.

Although the entire speech record is watermarked in Ruiz's work [2], only selected frames of speech may be watermarked depending upon the requirements of the application (Figure 5). By watermarking selectively, a degree of unpredictability is introduced about the exact locations of the watermarks. It is preferable to watermark the higher intensity speech regions, since for a given CWR, the watermark intensity will be

greater in these regions. As a result, the embedded watermarks will be more robust against certain attacks. The results supporting this are presented in Chapter 4. The computational complexity and hence the amount of time required for watermarking will be reduced on watermarking selected speech regions.

2.2 Correlation detector

A correlation detector (type II detector, see Section 1.2) can be used to detect the presence of the watermark in the stego-signal when subjected to linear distortion. The detector uses the normalized correlation between the original and possibly distorted watermark recovery signals. The latter are obtained by taking differences between the respective stego-signals and the cover-signal. Such a detector may not appear to be necessary for the TEC strategy as the watermark recovery signal (possibly distorted) may be fed to the recovery process in any case. However, the correlation detector is useful for acquiring quantified information about the presence or absence of the watermark. This information is crucial, for example, when an attack hinders the recovery process. The correct alignment of the two watermark recovery signals (original and possibly distorted) is an essential requirement for the correlation detector to perform correctly.

If s' is a speech signal that differs from the original cover-signal by an added sequence η, the correlation detector can be used to obtain information about the presence or absence of the watermark in s' as follows.

$$s' = c + \eta. \tag{4}$$

The normalized correlation between $\hat{\tilde{w}}$ and $(s' - c)$ is defined as,

$$\rho = \frac{\hat{\tilde{w}} \cdot (s' - c)}{| \hat{\tilde{w}} | \, | (s' - c) |} \tag{5}$$

A high value of ρ indicates the presence of the watermark in s'.

If the distortion has the effect of misaligning the stego-signal and the original, then the samples must be resynchronized before using the correlation detector. Due to such a requirement the correlation detector can be used for studying the effectiveness of the algorithm for watermark recovery from cropped speech as described in the next chapter.

2.3 Security and robustness

A good watermarking technique is one for which the security relies on the key and not on the secrecy of the algorithm. Public knowledge of the watermarking technology must not compromise security. This holds true for TEC-based speech watermarking. Security means that only authorized parties can decode the watermark [26]. It entails unpredictability and non-invertibility [23]. Non-invertibility of the watermarking technique means, for a modulated or encrypted watermark signal, it is practically impossible to find a fake watermark that can be produced by the same process [23].

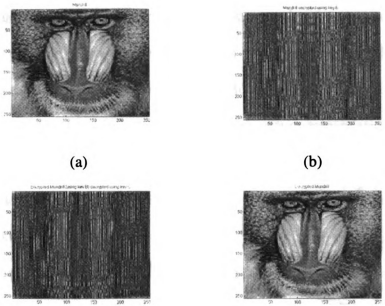
TEC speech watermarking derives its security from the quasi m-arrays used for cover-signal and watermark encryption. The recovery of the watermark is only possible with the knowledge of two quasi m-arrays (encryption keys) per frame. Further, encryption ensures secure transmission across the communication channel. It also helps in data access control i.e., an unauthorized person cannot retrieve the information [23]. Mere encryption without watermarking cannot provide copyright protection, as the data are unprotected and open to content tampering and copyright violation [23]. Hence it is important to hookup encryption and watermarking for secure copyright protection.

It is generally recommended [20] that the unmarked original not be publicly released. Enhanced security is also achieved by embedding the watermarks in random locations of the stego-signal rather than predictably throughout the entire stego-signal. Further, different signals being watermarked differently, copies of the same signal similarly watermarked (alternately, better to have copies of the stego-signal rather than the cover-signal), having more than one watermark (preferably different watermarks and keys) in a particular stego-signal, and using keys of different dimensions, all can contribute to security. The use of different keys avoids the obsolescence of the watermark if a set of keys used for watermark recovery were by chance made public knowledge after intentional tampering or copyright violation. Using quasi m-arrays and gold code arrays (keys) of higher dimension achieves greater encryption security. This is because, the number of available quasi m-arrays or gold code arrays increases with their dimension. Greater security also implies increased computational complexity implying a trade-off involved between increased security and computational burden.

TEC's masking algorithm provides additional protection by using different parameters (stego-keys) in a random fashion while ensuring the imperceptibility of the watermark.

The amount of data, measured in bits, needed to embed one unit of watermark information is termed the *watermark granularity* [17]. Finer granularity may result in greater robustness against certain attacks. In the case of cropping, for example, spreading the watermark across a large number of cover-signal samples implies greater risk of sample loss. However, finer granularity works against higher key dimension and hence security.

For the robustness of the watermarking scheme, the CWR plays a very crucial role. Lower CWR contributes to increased robustness. However the need for a perceptually transparent watermark places a practical lower bound on the CWR.



(a) (b)

(c) (d)

Figure 6. Encryption and decryption processes. (a) Original "mandrill" image. (b) Mandrill encrypted using key B. (c) Decryption of the encrypted mandrill in (b) using key C. If a signal is encrypted using key A (key B), it can be decrypted only using key A (key B). By using a different key for decryption, the mandrill gets encrypted twice. (d) Decryption of the encrypted mandrill in (b) using key B.

The next chapter focuses on the robustness of TEC-based speech watermarking to additive noise, cropping and protocol attacks.

Chapter 3

ROBUSTNESS STUDY

The issue of watermark robustness is introduced in Chapter 1. Robustness is the ability of the watermark to survive attacks and other forms of distortion. A watermarking scheme is said to be robust against a particular attack, if watermark detection and recovery are possible. This chapter deals with the robustness of TEC-based speech watermarking to additive noise, cropping and protocol attacks.

The encryption capabilities of TEC ensure secure transmission of stego-signal across a communication channel and secure storage in an archive. If a hacker tries to intercept (or download) and then attempts to decrypt a transmitted (or stored) stego-signal without the knowledge of the encryption keys, the nominal stego-signal obtained on decryption will be unintelligible. Hence, it is sufficient and necessary to use the unencrypted stego-signal for robustness study. According to [3], the TEC encrypted signal is insensitive and robust to channel noise. Due to the all pass nature of the encryption and decryption processes, the noise strength in every sample of the decrypted signal will be small.

The main focus of this chapter is robustness to cropping and additive noise. Cropping results in irretrievable loss of information that causes desynchronization in the watermark detection and recovery processes. As a consequence, the watermark fails to be detected and recovered. A number of watermarking techniques, especially signal (spatial or time) domain techniques are vulnerable to the damage caused by cropping. The damage caused by cropping depends more on the watermark embedding (for example, according to an additive rule) or detection strategy, than on the nature of the watermark.

In order to tackle cropping, an algorithm for watermark detection and recovery from cropped speech is presented. This algorithm can be applied to any cropped stego-signal, even if watermarked using a different watermarking technique.

3.1 Additive noise

Addition of uncorrelated and randomly generated noise is a common attack against a watermarked stego-signal. Techniques where the watermark is in the form of LSB modifications, are especially prone to such an attack or distortion. To study the robustness of TEC speech watermarking to additive noise, independent, uncorrelated and randomly generated noise was added to every sample of the stego-signal. The noise amplitude was either uniformly distributed or Gaussian distributed as shown in Figure 7. If s' is the noisy stego-signal, then

$$s' = c + \tilde{\tilde{w}} + \eta \tag{6}$$

The recovered watermark signal will now be,

$$\hat{\tilde{\tilde{w}}}' = \tilde{\tilde{w}} + \eta \tag{7}$$

As implied by equation (6), the robustness of the watermarking technique to additive noise depends upon the watermark to noise power ratio (WNR). The significance of a particular value of the WNR cannot be ascertained independently of the following factors:

i) Stego-signal to noise ratio (SNR), defined as

$$SNR_{dB} = 10\log_{10}\frac{P_s}{P_\eta} \tag{8}$$

where P_s and P_n are the signal and noise energy, averaged over the entire duration of the speech sequence (that is the signal and noise power).

$$P_s = \frac{1}{N}\Sigma s[n]^2 \tag{9}$$

$$P_\eta = \frac{1}{N}\Sigma \eta[n]^2 \tag{10}$$

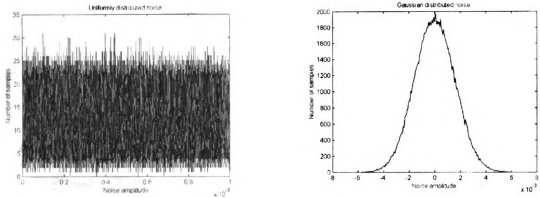
$s[n]$ and $\eta[n]$ are samples of the stego-signal and noise at time n.

ii) Cover-signal to watermark ratio (CWR[2]).

The CWR is influenced by whether a constant or an adaptive gain factor is used in the masking algorithm. If a constant gain factor (k factor), is used, then the CWR varies from samples to sample. On the other hand, when an adaptive gain factor ($k[n]$ factor) is used, the CWR is a constant throughout. In this case, robustness will also depend on the temporal placement of the watermark in the cover-signal. Since the intensity of the embedded watermark is adapted to the intensity of the cover-signal, the strength of the watermark will be greater in the higher intensity regions of speech. Experimental results for different CWRs, SNRs, and watermarks are presented in Chapter 4. To assess the significance of the experimental results, a group of individuals were asked to look at or listen to the results. The squared error (E) and normalized correlation (ρ) were used in conjunction, to provide quantified information. It was inferred from the experimental results that the damage survived by the watermark was sufficient to lower the commercial value of the attacked stego-signal, when the embedded watermark was mildly perceptible. A detailed discussion of the results is presented in Chapter 4.

---

[2] defined in (1) as, $CWR_{dB} = 10\log_{10}\frac{\tilde{C}[n]}{k[n]\times\tilde{W}[n]}$



(a) (b)

Figure 7. Noise amplitude distribution

3.2 Cropping

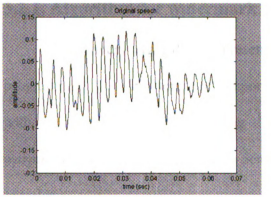
Cropping is an attack on the content of the stego-signal wherein samples of the signal are deleted in a random or deterministic manner. About 1 in 50 speech samples may be cropped without introducing any perceptible difference. Cropping may be an intentional attack or unintentionally introduced distortion. It is extremely easy to implement, but most digital watermarking schemes are vulnerable to the damage caused by it.

One method of identifying the attack to be cropping is by making use of the cross-correlation between the original watermark recovery signal (obtained by taking the difference between an undistorted stego-signal and the cover-signal) and the attacked watermark recovery signal. If samples are indeed cropped from the stego-signal, the normalized cross-correlation continues to sharply decrease as more and more cropped samples are encountered.

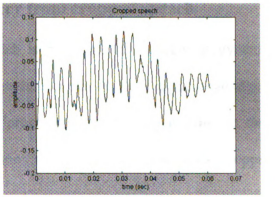
Cropping desynchronizes the recovery process, making watermark recovery difficult. Hence there is a need for an algorithm to identify the cropped samples, and to undo the damage caused by cropping, in order to make possible watermark recovery.

(a)

(b)



(c)



(d)

Figure 8. Cropping in images and speech. (a) Original mandrill image. (b) Cropped mandrill image. (c) 1000 samples from the speech "Theodore Roosevelt talks about Wilson and Taft". (d) A cropped version of the speech in (c). About 1 in 50 samples were cropped. It can be observed that there is greater predictability in the manner in which cropping manifests itself in images than in speech.

Duric *et al.* [22] make use of *registration patterns* (invariant features of an image) to recognize and restore images that are subjected to detection-disabling affine transformations. Typically, the registration patterns, also known as identification marks might be groups of points that exhibit uniqueness. Watermarks are then recovered from

the restored images. Such a methodology works for images, due to the manner in which cropping manifests itself in images. On cropping, the aspect ratio, shape or resolution of the image is generally affected. Hence, the effects of cropping are more predictable in the case of images (see Figure 8). Due to the random nature of speech, the geometry does not facilitate the derivation of registration patterns that are unique and invariant to transformations. Even in the case of images, the registration patterns can be exploited or attacked [20] to undermine their functionality.

Hence, in order to deal with cropping in speech, a dynamic programming based approach to identify the cropped samples and undo the damage was favored.

3.2.1. Algorithm for watermark recovery from cropped speech

A recovery algorithm is presented which is based on the concept of dynamic programming [5]. An attempt to temporally align the samples of the cropped stego-signal with the original stego-signal using dynamic programming (and hence *dynamic time warping* (DTW) [5]) will inherently determine the (former) time locations of the cropped samples.

Consider the i-j plane (as shown in Figure 9) with the cropped stego-signal (test string) along the i-axis and the stego-signal (reference string) along the j-axis. Determination of the cropped samples is treated as the problem of finding the minimum distance path through the grid. A path is a collection of nodes of the form $(t(i), s(j))$ connecting the original and terminal nodes. Distances or costs are assigned to paths in the form of nodal costs. The cost associated with the node $(t(i), s(j))$ is defined as,

$$d_n(i, j) = (t(i) - s(j))^2. \tag{11}$$

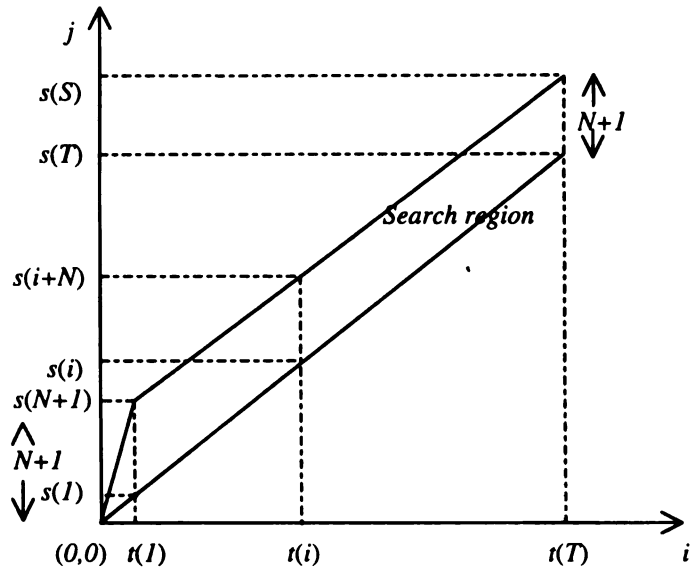


Figure 9. Dynamic programming approach to recovering cropped speech samples.

The search for the optimal path is described as follows. Let S be the length (total number of time samples) of the uncropped stego-signal, and T be the length of the cropped stego-signal. Assuming that no additional or duplicate samples are added to the stego-signal, the number of samples cropped is

$$N = S - T \,. \tag{12}$$

The following *search constraints* are imposed on the search region to limit the amount of computation and to ensure appropriate matching between the test and reference strings:

Monotonicity. For the path to be monotonic it must advance in the upward direction, i.e., it should not go "south" or "west" in the grid. Further, movement of the path in the horizontal or the vertical direction is prohibited as a single test sample cannot be associated with more than one reference sample and *vice versa*.

Global path constraints. Since N samples are cropped and the path can only move in the upward direction, element $t(i)$ of the cropped stego-signal can be matched only with the

($N+1$) elements $s(i)$ *to* $s(i+N)$ of the stego-signal. A similar constraint is applied at the endpoints. The result is a constrained search region in the form of a diagonal strip as shown in Figure 9.

Local path constraints. As every sample of the cropped stego-signal is contained in the original stego-signal, the optimal path should include *all* the test string elements. That is, no skips are permitted along the i-axis. At most, N reference string samples may be skipped in the process of finding the optimal path, as N samples were cropped. Thus, for node ($t(i)$, $s(j)$) in the search region, the possible immediate predecessor nodes include ($t(i-1)$, $s(k)$) where k ranges from ($i-1$) to ($j-1$).

As a consequence of the *Bellman optimality principle* [5], the optimal path to the node ($t(i)$, $s(j)$) can be found by considering the best paths associated with all the possible predecessor nodes and choosing the one with the minimum cost,

$$D_{min}(i, j) = \min_{(i-1, k)}\{D_{min}(i-1, k) + d_n(i, j)\},\ k = (i-1),\ldots,(j-1) \qquad (13)$$

After all the nodes in the search region are considered, a set of $N+1$ optimal paths is obtained. The first path, that is the one that involves zero skips, is the same as the cropped stego-signal. The global optimal path is the one associated with least cost among them. If the first path is associated with the least cost, then it implies that the last N samples of the stego-signal were cropped. It can be observed that the number of optimal paths is one more than the number of cropped samples. This follows as a direct consequence of the search constraints and equation (13). Although the paths might have common nodes, they never traverse each other. At every node ($t(i)$, $s(j)$) of a particular optimal path, it is necessary to record the immediate predecessor node from which the

path was extended. This way the path may be reconstructed by backtracking beginning at the terminal node.

The overall algorithm based on the principles above involves the following steps:

i) *Initialization:* The original node is $(0,0)$ and the nodal cost associated with it is zero. $(0,0)$ is the only predecessor associated with nodes $(t(1), s(j))$, j = 1,...,$(1+N)$.

$D_{min}(1, j) = d_n(0,0) + d_n(1, j), \quad j = 1,....,(1+N)$

$\psi(1, j) = (0,0), \quad j = 1,....,(1+N)$

$\psi(1, j) =$ *the index of the predecessor node to* $(1, j)$.

$\delta_1(j) = D_{min}(1, j), \quad j = 1,....,(1+N)$

ii) *Recursion:*

For $i = 2,...,T$

For $j = i,...,(i+N)$

Compute $D_{min}(i, j)$ using (13).

($D_{min}(i-1, j)$ is held in $\delta_1(j)$).

($\psi(1, j)$ is recorded for every (i, j)).

$\delta_1(j) = D_{min}(i, j)$

Next j

Next i

iii) *Termination:* The best path is the one associated with the least cost.

min $\{D_{min}(T, j)\}, \quad j = T,...,(T+N)$

iv) *Reconstruction:* The best path accurately identifies samples of the cropped stego-signal that are present in the stego-signal. The cropped samples are the ones, which are not present in the cropped stego-signal. The reconstructed stego-signal can be obtained

easily by reinserting the cropped samples at the appropriate places of the cropped stego-signal.

v) *Watermark recovery:* The watermark recovery process is applied to the reconstructed stego-signal.

3.2.2. Memory and computational requirements

The algorithm requires about $(N+1)T$ nodal costs or distance measures to be computed and approximately $((N+1)(N+2)T)/2$ implementations of equation (13). Considering the memory requirements, a matrix of size $O(TS)$ must be allocated for backtracking. This requirement cannot be replaced by the use of $N+1$ arrays of size $O(T)$ each. Such a replacement will require precise knowledge of the nodes comprising each path and this information will not be available until the entire algorithm has been executed. To compute $D_{min}(i, j)$ at every (i, j) within the search region, it is necessary to have just the past $D_{min}(i-1, j)$ values for $j = (i-1),...,(j-1)$. Therefore, at most an array of dimension $1\times(N+1)$ is required assuming that the computation can be done in-place.

3.2.3 Cropping in the presence of additive noise

Though TEC speech watermarking is fairly robust to additive noise, the recovery process is severely affected even if one sample is cropped. However, it was found that the DTW algorithm for watermark detection and recovery functioned quite efficiently in the presence of independent uncorrelated random noise. The experimental results for different SNR and CWR values are discussed in the next chapter.

3.3 Counterfeit attacks

Also known as *protocol attacks* [11, 18, 21, 24, 25], *counterfeit attacks* seek to undermine the concept of watermarking itself by producing fake originals or fake watermarked signals. Counterfeit attacks are not concerned with destroying the embedded watermark nor disabling the recovery process. In the context of counterfeit attacks, robustness has a different meaning. A watermarking scheme is said to be *robust against counterfeit attacks* if the attack does not succeed in creating ambiguity in the resolution of ownership (or any other purpose for which watermarking is used).

There are different types of counterfeit attacks including *inversion attacks, multiple watermarks*, and *copy attacks*. The basic idea behind watermark copy attack is to copy a watermark from a stego-signal to another signal without the knowledge of the watermarking algorithm and the key that were used to create the rightful stego-signal [15, 25]. This is achieved by estimating the embedded watermark either by direct prediction or denoising [25]. In the case of TEC speech watermarking, the coefficients of the embedded doubly encrypted watermark are outcomes of Gaussian random variables. For watermark recovery, a good estimate of these coefficients and knowledge of encryption keys will be essential. Thus, the copy attack will be extremely difficult to implement in the case of TEC watermarking.

In an inversion attack, the attacker subtracts his or her watermark from the stego-signal. The attacker thus obtains a fake cover-signal (original) and claims to be the owner of the watermarked signal. This can create ambiguity in the resolution of the ownership of the stego-signal. Craver *et al.* [11] show that non-invertibility of the embedded

watermarks is essential for robustness against inversion attack. Non-invertibility of TEC speech watermarking is discussed in Section 2.3.

The problem of multiple watermarks arises when an attacker inserts another watermark into the already watermarked signal and claims ownership of the signal. As a consequence, this creates ambiguity in the resolution of ownership. The TEC speech watermarking technique can be made robust against such a problem as discussed below.

Suppose person A is the real owner of a speech watermarked using TEC. A's stego-signal is,

$$s = c + \tilde{\tilde{w}} \tag{14}.$$

Person A releases only the watermarked speech and not the cover-signal to the public. Person B obtains a copy of s and is interested in selling illegal copies. B embeds another watermark w_B into s and circulates illegal copies of s_B. It is assumed that w_B is embedded in s in accordance with an additive rule and also that w_B is not correlated with s. This ensures that the distortion produced as a result of watermarking the already marked (using TEC) stego-signal is linear and uncorrelated. Robustness against distortion that is non-linear and correlated with the stego-signal is beyond the scope of this research.

$$s_B = c + \tilde{\tilde{w}} + w_B \tag{15}.$$

Person A comes across one of the illegal copies and recovers $\tilde{\tilde{w}}$ from it. When A tries to sue B, B claims ownership of s_B. However this fails to create enough ambiguity due to the following reasons.

i) It will not be possible for B to show a copy of the speech that does not contain A's watermark.

ii) A has the cover-signal and stego-signal that do not contain B's watermark, giving credence to the proposition that A is the true owner.

iii) Copies of the stego-signal (if any) not circulated by B contain A's and not B's watermark.

Chapter 4

IMPLEMENTATION DETAILS, RESULTS AND CONCLUSIONS

4.1 Robustness testing engine

In this chapter, the experimental results obtained by testing the robustness of TEC watermarking to additive noise and to cropping are presented.

One main problem faced by the current digital watermarking technology is the absence of common benchmarks for the evaluation of different watermarking schemes. Petitcolas [28] proposes the establishment of a public benchmarking service. The performance metrics to be used for evaluation are yet to be established. Software packages StirMark [29] and unZign [30] include robustness testing engines for image watermarks. Such services provide a common platform for the evaluation of different watermarking techniques. Public domain software for testing the robustness of audio watermarking techniques is not yet available.

To study the robustness of TEC speech watermarking, the robustness testing engine developed by Ruiz *et al.* [35] was used. The testing engine can be used to perform 17 tests on the stego-signal. The tests include addition of random noise, cropping, filtering, μ-law compression and expansion. The robustness testing engine accommodates a high degree of flexibility for setting the parametric values characterizing the tests. For the evaluation of TEC speech watermarking, an error measure was determined according to the following equation.

$$E = \frac{(\hat{w} - w')^2}{\hat{w}^2} \tag{16}$$

In (16), $\hat{w}$ indicates the original watermark recovery signal and w' indicates the watermark recovery signal obtained from a distorted stego-signal. In addition to the error E, the normalized correlation ρ, defined in (5), is used to evaluate the performance. A quality rating (see Table 1) on a scale from 1 to 5 was used to quantitatively describe the perceived results. Martin used a similar rating in [34] to rank the quality of the watermarked image. Two individuals were asked to rate the quality of the stego-signal, distorted stego-signal and the watermarks recovered form the distorted stego-signal in accordance with Table 1. These ratings were obtained without providing the individuals with the knowledge of the error and normalized correlation values. A quality rating of 3 for the recovered watermark is considered sufficient and it implies that the watermark is identifiable.

Table 1 – Quality rating

Rating	Quality of the watermarked signal	Quality of the recovered watermark	Effect of distortion on the stego-signal
1	Watermark imperceptible	Excellent	No perceptible damage
2	Perceptible, not annoying	Good	Perceptible
3	Slightly annoying	Fair	Mildly degrading
4	Disturbing	Poor	Degrading
5	Very disturbing	Bad	Destructive

4.2 Robustness to additive noise

Experiments were performed for studying the robustness of the stego-signal against uncorrelated additive noise. Gaussian or uniformly distributed noise was used. For all the experimental results and simulations presented in this section, a record of 48387 samples (3 seconds) of the speech "Theodore Roosevelt talks about Wilson and

Taft" [32] was used as source material. The signal is monaural, sampled at 16kHz with 16-bit quantization. The "Lena" image was used as the watermark.

Table 2 enumerates the experimental results obtained by adding randomly generated Gaussian noise to the stego-signal. A set of three watermarks was embedded in the 48387-sample speech waveform, by dividing it into three frames, each consisting of 16129 samples (Figure 10). For the last five entries in the table, a watermark was embedded selectively in *the first frame* as it was associated with higher speech energy. For all results in Table 2, the masking algorithm used of a constant gain factor. Since every sample of the encrypted watermark was scaled by a constant, the CWR as defined in (1) was not a constant, but a varying quantity. In Table 2 the average CWR values over every frame and across the entire speech segment are shown. When a constant gain factor is used, the mean CWR across the entire speech segment varies widely from the CWRs averaged across individual frames. The SNR and the normalized correlation between the distorted and original watermark recovery signals are tabulated.

It can be inferred from Table 2 that robustness against Gaussian additive noise depends on the CWR and the SNR. A lower CWR and a higher SNR contribute to increased robustness. Since the embedding process is independent of the speech intensity, watermarking selectively does not contribute to increased robustness and all the recovered watermarks are of the same quality. In interpreting the normalized correlation value, it must be noted that it is dependent on both the SNR and CWR. Even if the SNR is low, the normalized correlation between the original and distorted watermark recovery signals may be high, if the CWR is low. When the embedded watermarks were very mildly perceptible, corresponding to a mean CWR of approximately 26 dB, the mean of

the recovered watermarks was identifiable for an SNR of 42.63dB. In this case, the noise had the effect of just mildly degrading the stego-signal. When the mean CWR was 21.4 dB, better robustness was exhibited. However, the watermark was perceptible in the speech.

Table 2 – Robustness to Gaussian noise (constant gain factor)

CWR (dB)			Mean CWR (dB)	Gaussian noise		SNR (dB)	Norm. Correlation ρ	Quality of stego-signal	Effect of noise on stego-signal	Recovered watermarks			
1	2	3		μ	σ					M	1	2	3
31.4	19.1	27.6	26.04	0	.0046	26.61	0.3491	2	4	5	5	5	5
31.4	19.1	27.6	26.03	0	.00092	42.63	0.8784	2	2	3	4	4	4
31.4	19.1	27.6	26.06	0	.00009	62.60	0.9985	2	2	2	2	2	2
26.4	14.2	22.6	21.06	0	.0092	22.59	0.3161	2	4	5	5	5	5
26.4	14.2	22.6	21.05	0	.0046	28.63	0.5479	2	3	4	5	5	5
26.4	14.2	22.1	21.06	0	.0037	30.59	0.6302	2	3	3	4	4	4
26.4	14.2	22.6	21.06	0	.0023	34.67	0.7925	2	3	3	4	4	4
26.4	14.2	22.6	21.06	0	.00092	42.64	0.9562	2	2	2	2	2	2
21.4	9.2	17.6	16.06	0	.0046	28.68	0.7605	3	4	3	4	4	4
26.4	-	-	26.38	0	.0092	22.55	0.1865	2	4	5	5	-	-
26.4	-	-	26.38	0	.0023	34.65	0.5988	2	3	4	4	-	-
26.4	-	-	26.38	0	.00092	42.58	0.8805	2	2	3	3	-	-
21.4	-	-	21.38	0	.0046	28.60	0.5531	3	4	4	4	-	-
21.4	-	-	21.38	0	.00092	42.63	0.9581	3	2	2	2	-	-

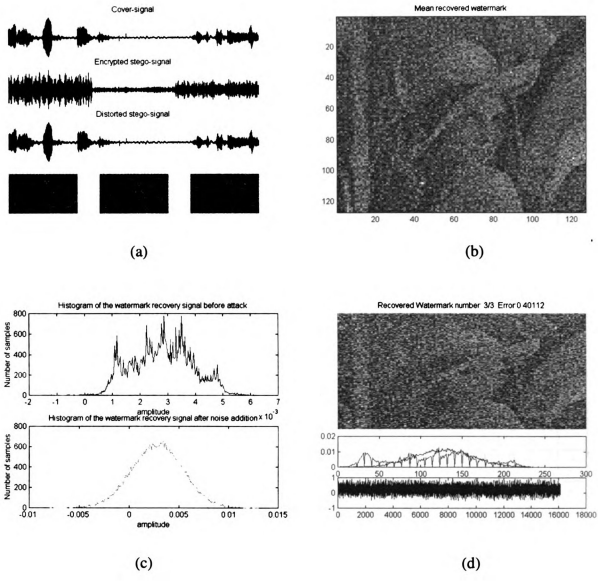


Figure 10. Robustness of TEC watermarking to Gaussian noise (constant gain factor). 48387 samples of the speech "Theodore Roosevelt talks about Wilson and Taft" [32] was used as the cover-signal. The "Lena" image was used as the watermark. (a) The cover-signal, encrypted stego-signal, the stego-signal distorted by the addition of Gaussian noise and the watermarks recovered from the distorted stego-signal. The recovered watermarks were associated with a quality rating of 4 (Table 2). (b) Mean recovered watermark (quality rating of 3) (c) Histogram of the watermark recovery signal before and after the addition of Gaussian noise. (d) One of the recovered watermarks. Also shown are the histograms of the original and recovered watermarks, and the encrypted watermark reshaped into an array.

Table 3 – Robustness to Gaussian noise (adaptive gain factor)

CWR (dB)	Gaussian noise		SNR (dB)	Error (E)			Norm. Correlation ρ	Quality of stego-signal	Effect of noise on stego-signal	Recovered watermarks			
	μ	σ		1	2	3				M	1	2	3
23.90	0	.0009	42.59	.071	.6202	.203	0.9599	3	2	3	3	5	3
29.90	0	.0046	28.70	.784	1.076	.957	0.3275	1	4	5	5	5	5
29.90	0	.0009	42.66	.203	.8418	.433	0.8657	1	2	3	3	5	4
29.90	0	.0001	62.67	.002	.0754	.007	0.9983	1	1	2	1	3	2
26.91	0	.0046	28.64	.703	1.087	.869	0.4396	2	4	5	4	5	5
26.90	0	.0009	42.64	.115	.7401	.291	0.9253	2	2	3	3	5	4
22.04	0	.0009	42.66	.058	-	-	0.9500	3	2	3	3	-	-
26.04	0	.007	25.09	.849	-	-	0.2435	2	4	5	5	-	-
26.91	0	.0069	25.13	.959	-	-	0.3071	2	4	5	5	-	-
26.05	0	.0046	28.65	.709	-	-	0.3541	2	4	5	5	-	-
26.05	0	.0009	42.65	.135	-	-	0.8873	2	2	3	3	-	-
25.05	0	.0009	42.65	.119	-	-	0.9061	2	2	3	3	-	-
28.05	0	.0009	42.63	.179	-	-	0.8364	1	3	3	3	-	-
28.04	0	.0001	62.68	.002	-	-	0.9979	1	1	2	2	-	-

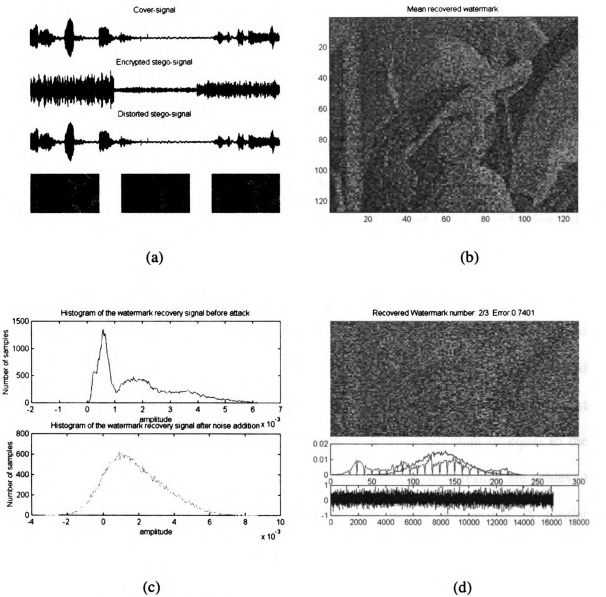


Figure 11. Robustness of TEC watermarking to Gaussian noise (adaptive gain factor). 48387 samples of the speech "Theodore Roosevelt talks about Wilson and Taft" [32] was used as the cover-signal. The "Lena" image was used as the watermark. (a) The cover-signal, encrypted stego-signal, the stego-signal distorted by the addition of Gaussian noise and the watermarks recovered from the distorted stego-signal. The recovered watermarks were associated with a quality ratings of 3, 5 and 4 respectively (Table 3). (b) Mean recovered watermark (quality rating of 3) (c) Histogram of the watermark recovery signal before and after the addition of Gaussian noise. The shape of the histogram before attack indicates the use of an adaptive gain factor for watermark embedding (d) One of the recovered watermarks. Also shown are the histograms of the original and recovered watermarks, and the encrypted watermark reshaped into an array.

Table 3 shows the results obtained by testing the robustness of TEC speech watermarking against Gaussian noise when the masking process uses an adaptive gain factor (also see Figure 11). Hence, the CWR is a constant throughout the speech. The error, determined according to (16) is also tabulated. In addition to the CWR and the SNR, robustness (and the error) is influenced by the intensity of the speech. On comparing Tables 2 and 3, it is inferred that for a given quality of the watermarking process, better robustness is exhibited by speech watermarked using an adaptive gain factor. This fact suggests the use of masking algorithms that exploit the perceptual properties of human auditory system for increased robustness. The ultimate aim would be to achieve robustness such that the quality of the recovered watermarks is better than, or comparable to, the effect of noise on the stego-signal, even when the embedded watermarks are imperceptible. That is, the rating in the recovered watermarks or at least the mean recovered watermark column (see Table 2 or 3) is less than or equal to the rating in the effect of noise on stego-signal column. At present, such results are achieved only for CWRs that cause watermarks to be at least mildly perceptible. A similar behavior was observed when experiments were conducted using non-zero mean Gaussian noise.

Experiments were conducted to study the robustness of TEC watermarking to uniformly distributed noise (see Figure 12). The results are tabulated in Tables 4 and 5. When the embedded watermarks were mildly perceptible, at least one of the recovered watermarks was identifiable for an SNR of approximately 30dB for the adaptive gain factor case. When the CWR was 31.4 dB, the mean recovered watermark was identifiable

for an SNR of 28.34 dB. Taking into account the CWR, robustness of TEC watermarking tends to be better in the presence of uniformly distributed noise over Gaussian noise.

Table 4 – Robustness to uniformly distributed noise (constant gain factor)

CWR (dB)			Mean CWR (dB)	Noise (uniform)		SNR (dB)	Norm. Correlation ρ	Quality of stego-signal	Effect of noise on stego-signal	Recovered watermarks			
1	2	3		μ	max.					M	1	2	3
31.4	19.1	27.6	26.0	.0046	.0092	27.40	0.8876	2	4	5	5	5	5
31.4	19.1	27.6	26.0	.0023	.0046	33.41	0.9259	2	3	4	5	5	5
31.4	19.1	27.6	26.1	.0012	.0023	39.43	0.9625	2	2	3	3	3	3
26.4	14.1	22.6	21.1	.0041	.0083	28.34	0.9251	2	3	3	4	4	4
26.4	14.1	22.6	21.0	.0023	.0046	33.39	0.9568	2	3	3	3	3	3
21.4	9.16	17.6	16.1	.0046	.0093	27.38	0.9512	3	4	3	4	4	4
21.4	9.15	17.6	16.1	.0035	.0070	29.09	0.9647	3	3	3	3	3	3

Table 5 – Robustness to uniformly distributed noise (adaptive gain factor)

CWR (dB)	Noise (uniform)		SNR (dB)	Error			Norm. Correlation ρ	Quality of stego-signal	Effect of noise on stego-signal	Recovered watermarks			
	μ	max		1	2	3				M	1	2	3
29.89	.0046	.0092	27.39	.178	.294	.233	0.8103	1	4	5	5	5	5
29.91	.0035	.0069	29.90	.141	.286	.209	0.8365	1	3	5	4	5	5
29.90	.0012	.0023	39.44	.046	.214	.099	0.9351	1	2	3	3	5	3
26.90	.0035	.0069	29.92	.119	.271	.172	0.8682	2	4	4	4	5	4
23.92	.0042	.0083	28.29	.111	.259	.157	0.8838	3	4	4	3	5	5
23.90	.0035	.0069	29.90	.077	.248	.138	0.9006	3	3	3	3	5	4

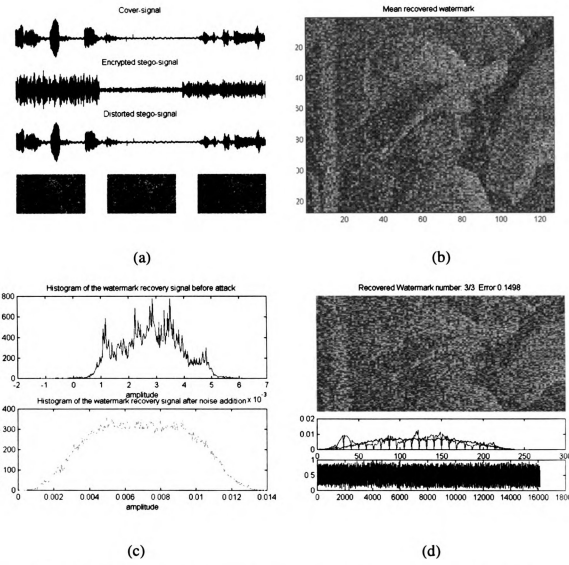


Figure 12. Robustness of TEC watermarking to uniformly distributed noise (constant gain factor). 48387 samples of the speech "Theodore Roosevelt talks about Wilson and Taft" [32] was used as the cover-signal. The "Lena" image was used as the watermark. (a) The cover-signal, encrypted stego-signal, the stego-signal distorted by the addition of noise and the watermarks recovered from the distorted stego-signal. The recovered watermarks were associated with a quality ratings of 4 (Table 4). (b) Mean recovered watermark (quality rating of 3) (c) Histogram of the watermark recovery signal before and after the addition of Gaussian noise. (d) One of the recovered watermarks. Also shown are the histograms of the original and recovered watermarks, and the encrypted watermark reshaped into an array.

4.3 Robustness to cropping

The DP algorithm for the detection of cropped speech samples and watermark recovery is described in Section 3.2.1. The Matlab implementation differs slightly from the description found in 3.2.1. This deviation was necessary to account for the "out of memory " problems encountered in Matlab when the algorithm was used for a speech sequence consisting of more than approximately 7000 samples.

4.3.1. Implementation details of the modified DP algorithm

The algorithm described in Chapter 3 requires a matrix of size $O(TS)$, where T and S are the lengths of the cropped and original stego-signals, respectively. The values of T and S employed here result in out of memory problems when the unaltered DP algorithm is implemented in Matlab. One simple remedy would be to break down the long speech sequence (greater than 7000 samples) to shorter sequences and to apply the algorithm separately to each of them. However, this would necessitate the determination of the exact number of cropped samples in each of the shorter segments. For this, the exact end-points may have to be determined by cross-correlation between the original stego-signal and cropped speech segment in the appropriate region. Such an approach may not perform optimally in the presence of noise, as noise might hinder the accurate determination of the end points.

Hence, the implementation of the algorithm was modified to alleviate out of memory problems or the need to determine the exact number of cropped samples in each of the shorter speech segments. The modified form determines the global best path and requires 'p' matrices, each of which comprises of m rows and $m+N$ columns. Here, m is a

number less than 8000 and greater than N, the number of cropped samples. The modification involves dividing the cropped stego-signal into frames of m samples each except the last one, which may contain less than m samples. p is the total number of frames, *excluding* the last one. The search constraints described in Section 3.2.1 are applied here (Figure 13). The algorithm proceeds similar to the original version by assigning costs to nodes and applying the Bellman optimality principle. During transition from one frame to another, the costs associated with the last $N+1$ nodes [in the case of the first frame they include nodes $(t(m), s(j))$, $m \leq j \geq m+N$] of the previous frame are taken as the initial costs associated with the next frame. Backtracking information for each of the m-segment frames is stored in p matrices of dimension $(m, m+N)$. At the end of the last (that is, $p+1^{st}$) frame, the global best path is chosen from the $N+1$ optimal paths, by selecting the one with the least cost. On backtracking across the various frames, the global best path is reconstructed.

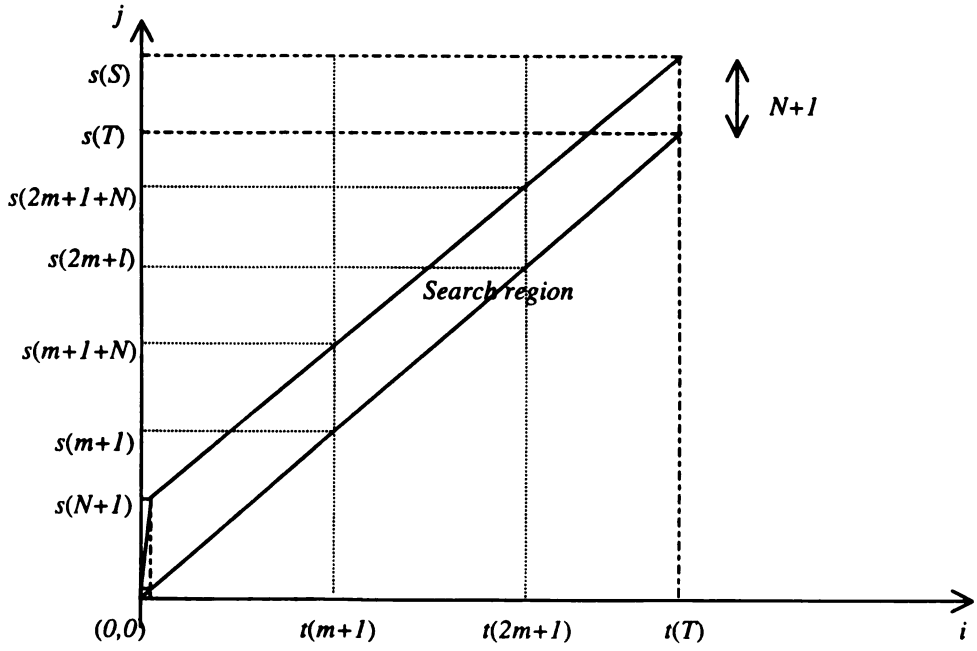


Figure 13. Modified implementation of the DP algorithm

The modified algorithm involves the following steps:

i) *Initialization:* The original node is $(0,0)$ and the nodal cost associated with it is zero. $(0,0)$ is the only predecessor associated with nodes $(t(1), s(j))$, j = 1,...,$(1+N)$.

$D_{min}(0, 0) = d_n(0,0)$, $i = j = 0$.

$\delta_1(j) = D_{min}(0, 0)$

ii) *Recursion:*

For $k = 1,...,p$

For $i = i+1,...,km$

For $j = i,...,(i+N)$

$D_{min}(i, j) = \min_{(i-1, j)}\{D_{min}(i-1, j) + d_n(i, j)\}$, $k = (i-1),....,(j-1)$

($D_{min}(i-1, j)$ is held in $\delta_1(j)$).

Record $\psi_k(i, j)$.

$\psi_k(i, j)$ = the index of the predecessor node to (i, j) in the k^{th} frame.

$\delta_1(j) = D_{min}(i, j)$

Next j

Next i

Next k

iii) *Termination:*

For $i = km+1,...,T$

For $j = i+1,...,S$

$D_{min}(i, j) = \min_{(i-1, j)}\{D_{min}(i-1, j) + d_n(i, j)\}$, $k = (i-1),....,(j-1)$

($D_{min}(i-1, j)$ is held in $\delta_1(j)$).

Record $\psi_{p+1}(i, j)$.

$\psi_{p+1}(i, j)$ = the index of the predecessor node to (i, j) in the last frame.

$\delta_1(j) = D_{\min}(i, j)$

Next j

Next i

The best path is the one associated with the least cost.

min $\{D_{\min}(T, j)\}$, $j = T,...,(T+N)$

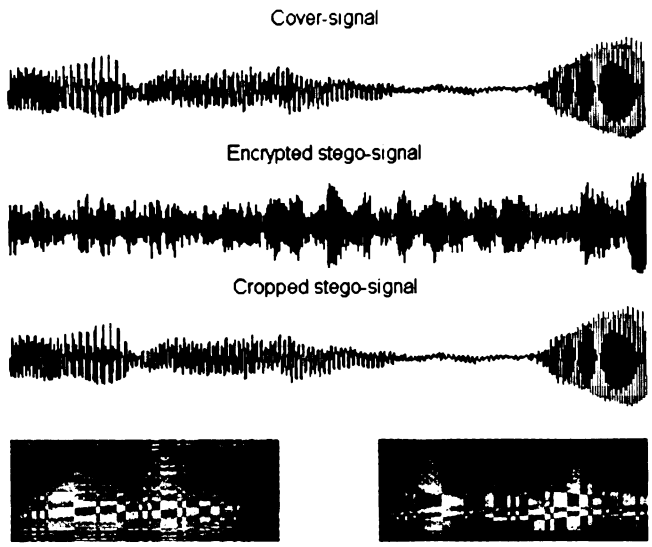
iv) *Reconstruction:* By backtracking through the $(p+1)$ frames, the global best path is obtained. The best path accurately identifies samples of the cropped stego-signal that are present in the stego-signal. The cropped samples are the ones, which are not present in the cropped stego-signal. The reconstructed stego-signal can be obtained easily by reinserting the cropped samples at the appropriate places of the cropped stego-signal.

v) *Watermark recovery:* The watermark recovery process is applied to the reconstructed stego-signal.

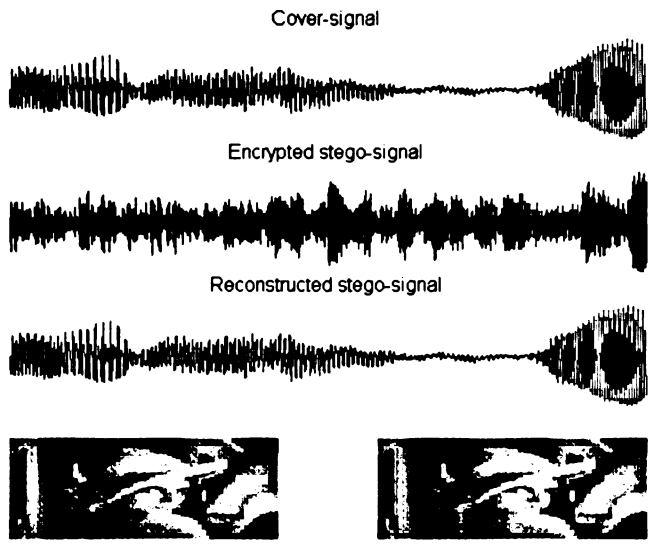
Computational requirements for this modified implementation are the same as those for the original algorithm of Section 3.2.2. Instead of a single matrix of size $O(TS)$, the modified implementation requires p matrices of size $O(m^2+mN)$, where m is small compared to T. This modification solves the out of memory problems.

4.3.2. Experimental results

As an example, the DP algorithm was applied to a cropped stego-signal watermarked using TEC. The cover-signal was obtained from the TIMIT speech database [33] and has a male voice saying: "She had your dark suit in greasy wash water all year." In Figure 12, 7968 speech samples were used. Quasi m-arrays of dimension 63×63 were used for encryption.



(a)



(b)

Figure 14. DP algorithm for watermark recovery (a) Cropping and watermarks recovered from cropped speech. (b) Reconstructed stego-signal obtained after the application of the DP algorithm. Watermarks recovered from the reconstructed stego-signal.

After the cover-signal was TEC watermarked, 150 samples of the stego-signal were randomly cropped using the robustness testing engine. The watermarks recovered from the cropped stego-signal are shown in Figure 14(a). The DP algorithm was then applied to the cropped stego-signal to detect the cropped samples and to reconstruct the stego-signal. The cropped samples were accurately determined and the watermarks recovered from the reconstructed stego-signal (Figure 14(b)).

The robustness of the DP algorithm was tested with varying CWRs and varying numbers of cropped samples. In the absence of additive noise, the cropped samples were accurately determined under all tested conditions .

Table 6 – Robustness to cropping and additive noise (adaptive gain factor)

CWR (dB)	Gaussian Noise		SNR (dB)	Number of cropped samples	Error (E)	Normalized correlation ρ	Cropped samples accurately determined Yes/No
	μ	σ					
32.2197	0	2.546×10^{-4}	44.9662	19	0.2663	0.9134	Yes
32.1181	0	0.0013	30.8328	19	0.8355	0.4365	Yes
32.2246	0	0.0023	25.7200	19	0.9234	0.3252	Yes
31.9115	0	0.0026	25.3634	19	0.9902	0.2323	Yes
32.0573	0	0.0025	25.2290	92	1.0290	0.1167	Yes
31.9341	0	0.0128	11.3201	3	1.2426	0.0391	Yes
32.2893	0	0.0115	12.464	19	1.1785	0.0296	Yes
31.8826	0	0.0115	12.4326	92	1.2323	0.0861	No
26.0974	0	0.0115	12.2692	92	1.2043	0.0960	No

4.4. Robustness to cropping in the presence of noise

In the previous section, it was determined that in the absence of noise, cropped samples were accurately determined. The DP algorithm was also tested for watermark recovery from stego-signals distorted by additive noise as well as cropping.. TEC speech watermarking was found to be fairly (Section 4.3) robust to additive noise alone when the embedded watermarks were not perfectly imperceptible. It is important for DP algorithm to tolerate noise, at least to the extent to which TEC speech watermarking is robust against additive noise.

Using Ruiz's robustness testing engine, the stego-signal was randomly cropped and subjected to additive noise. In all the experiments (Table 6), 961 samples of the utterance, "She had your dark suit in greasy wash water all year" [33] was used. The accuracy of the DP algorithm was verified by comparing the actual cropped samples with the missing samples detected by the algorithm. The performance is mainly dependent on the SNR. The algorithm is robust for a SNR of 11.3 dB or above. It was observed that when the SNR approaches the 11.3dB threshold, the performance degrades, with an increase in the number of cropped samples (see Table 6).

The DP algorithm is robust to additive noise well above the robustness threshold (approximately 30dB when the embedded watermarks were mildly perceptible) of TEC speech watermarking. Experiments have confirmed that for the range of importance, that is when the recovered watermarks are identifiable, the algorithm is reliable.

4.5. Conclusions

The salient points from the experiments described above are summarized as follows:

- The robustness of TEC speech watermarking to additive noise, is mainly dependent on the SNR and CWR. Higher SNRs and lower CWRs contribute to increased robustness. The need to maintain the perceptual transparency of the embedded watermark imposes a lower limit on the CWR.
- When the watermark masking algorithm involves the use of an adaptive gain factor, better robustness is exhibited by watermarks embedded in the higher intensity regions of speech.
- The DP algorithm for the detection of cropped samples and subsequent watermark recovery performs with 100% accuracy in the absence of noise. In the absence of noise, the performance is independent of the number of samples cropped.
- In the presence of cropping and uncorrelated additive noise, the performance of the DP algorithm is mainly determined by the SNR and the number of cropped samples.
- Unlike most watermarking techniques [9,10], TEC watermarking admits watermark recovery, not just watermark detection. If the watermark contains information supporting the owner or title, on recovery, this information will lead to greater credence in the true ownership.

Chapter 5

FUTURE WORK

Digital watermarking is an emerging technology and it faces problems typical of many new signal-processing endeavors. The main problems include difficulty in dealing with many types of attacks, a lack of standard tools with which to assess and compare watermarking schemes, and the lack of clear definitions of watermarking requirements [42]. As the need for the technology increases, these problems will have to be resolved if the methods are to be effective and therefore accepted by those depending on the technology for copyright protection. In addition to the challenges common to all watermarking techniques, TEC speech watermarking as described in this thesis, requires further research in a number of areas. Some areas identified for future work are as follows:

- Robustness of TEC watermarking to other attacks [45] must be studied. In particular, study of the robustness to signal-processing transformations like resampling, compression, filtering and quantization is of importance. These transformations may be the consequence of routine and unintentional operations on the stego-signal. Some of the other deliberate attacks to be studied include collusion attacks, cryptographic attacks and time-scale modification. While studying the robustness, it is also essential to test TEC watermarking against a combination of two or more attacks. Robustness study will entail developing a more elaborate robustness testing engine.
- The watermark-masking algorithm as described in this document, involves scaling the watermark by a gain factor in accordance with the CWR. Future work in this area comprises the application of masking algorithms that exploit the perceptual properties

of the human ear. Application of perceptual model-based masking algorithms becomes necessary for watermarking due to the rigid requirements of imperceptibility and robustness. Such an application must be viewed in conjunction with robustness against perceptual model-based compression algorithms like MPEG.

- Not much research has been done in the field towards understanding the embedding capacity [31] offered by the different watermarking techniques. Research must be done to determine the best strategies for utilizing the embedding capacity so as to fulfill watermarking and channel bandwidth requirements.
- An appropriate audio transform coding strategy must be implemented to effect compression in conjunction with watermarking. In such a scenario, it will be important to use audio or speech rather than image watermarks.
- In the context of various attacks (although this did not matter for additive noise and cropping), some classes of watermarks may perform than others. Future work will also pursue understanding watermark characteristics that result in optimum watermarks in the presence of a particular attack.
- TEC was initially developed as an image compression algorithm [3]. Application of TEC watermarking to images and the performance appraisal are areas in need of further work

REFERENCES

1. National Gallery of the Spoken Word, Michigan State U., http://www.ngsw.msu.edu

2. Fco. J. Ruiz and J.R. Deller, Jr., "Digital watermarking of speech signals for the National Gallery of the Spoken Word," *International Conference on Acoustics, Speech and Signal Processing 2000*, Istanbul, May 2000.

3. C.J. Kuo, J.R. Deller, Jr. and A.K. Jain, "Pre/post-filter performance improvement of transform coding," *Signal Processing: Image Communication*, vol. 8, pp.229-239, 1996.

4. C.J. Kuo and H.B. Rigas, "2-D quasi m-arrays and gold code arrays," *IEEE Transactions Information Theory*, vol. 37, pp. 385-388, March 1991.

5. J.R. Deller, Jr., J.H.L. Hansen, and J.G. Proakis, *Discrete Time Processing of Speech Signals* (2d ed.), New York: IEEE Press, 2000.

6. A. Gurijala and J.R. Deller, Jr., "Robust algorithm for watermark recovery from cropped speech," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing 2001*, Salt Lake City, May 2001 (in press).

7. J.R. Deller, Jr., A. Gurijala, and M.S. Seadle, "Audio watermarking techniques for the National Gallery of the Spoken Word," *Proc. 1st ACM-IEEE Joint Conference on Digital Libraries 2001*, Roanoke, Virginia, June 2001 (in press).

8. F.A.P. Petitcolas, R.J. Anderson and M.G. Kuhn, "Information Hiding – A Survey," *Proceedings of the IEEE*, special issue on protection of multimedia content, pp. 1062-1078, July 1999.

9. L. Boney, A.H. Tewfik and K.N. Hamdy, "Digital watermarks for audio signals," *Proc. IEEE International Conference on Multimedia Computing and Systems*, Hiroshima, pp. 473-480, June 1996.

10. P. Bassia and I. Pitas, "Robust audio watermarking in the time domain," *Proc. IX European Signal Processing Conference*, Rhodes, Greece, vol. I, pp.25-28, Sept. 1998.

11. M. Arnold, "Audio watermarking: features, applications and algorithms," *Proc. IEEE International Conference on Multimedia and Expo (II)*, pp. 1013-1016, 2000.

12. W. Bender, D. Gruhl, and N. Marimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol. 35, pp.313-336, 1996.

13. C.S. Lu, H.Y.M. Liao, and L.H. Chen, "Multipurpose Audio Watermarking", *Proc.15th International Conference on Pattern Recognition*, Barcelona, Spain, vol. III, pp. 286-289, Sept. 2000.

14. I.J. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Transactions on Image Processing*, vol. 6, pp. 1673-1687, Dec. 1997.

15. S. Voloshynovskiy, S. Pereira, T. Pun, J.K. Su, J.J. Eggers, "Attacks and Benchmarking, " submitted to *IEEE Communication Magazine*, 2001.

16. I.J. Cox, M.L. Miller, J. M.G. Linnartz, T. Kalker, "A review of watermarking principles and practices," *Digital Signal Processing for Multimedia Systems*, K. K. Parhi, T. Nishitani (eds.), New York: Marcell Dekker, Inc., pp. 461-485, 1999.

17. G.C. Langelaar, I. Setyawan, and R.L. Lagendijk, "Watermarking digital image and video data," *IEEE Signal Processing Magazine*, vol. 17, pp.20-46, Sept. 2000.

18. I.J. Cox, M.L. Miller, and J.A. Bloom, "Watermarking applications and their properties," *Proc. IEEE International Conference on Information Technology: Coding and Computing*, pp.6-10, Mar. 2000.

19. M. Wu and B. Liu, "Watermarking for image authentication," *Proc. IEEE International Conference on Image Processing*, Chicago, vol. 2, pp. 437-441, Oct. 1998.

20. I. J. Cox and J.P.Linnartz, "Some general methods for tampering with watermarks," *IEEE Journal on Selected Areas of Communication*, vol. 16, pp. 587-593, May 1998.

21. M. Ramkumar, A.N. Akansu, "Robust protocols for proving ownership of images," *Proc. IEEE International Conference on Information Technology: Coding and Computing*, Las Vegas, pp. 22-27, Mar. 2000.

22. Z. Duric, N.F. Johnson, and S. Jajodia, "Recovering watermarks from images," *Information and Software Engineering Technical Report*, ISE-TR-99-04, Apr. 1999.

23. G. Voyatzis and I. Pitas, "The use of watermarks in the protection of digital multimedia products," *Proceedings of the IEEE*, vol. 87, pp. 1197-1207, July 1999.

24. M. Ramkumar, A.N. Akansu, "Image watermarks and counterfeit attacks : Some problems and solutions", *Content Security and Data Hiding in Digital Media*, Newark, NJ, May 1999.

25. M. Kutter, S. Voloshynovskiy and A. Herrigel, "Watermark copy attack," *IS&T/SPIE's 12th Annual Symposium, Electronic Imaging 2000: Security and Watermarking of Multimedia Content II*, San Jose, vol. 3971, Jan. 2000.

26. J.K. Su, J.J. Eggers, and B. Girod, "Capacity of digital watermarks subjected to an optimal collusion attack," *Proc. European Signal Processing Conference*, Tampere, Finland, Sept. 2000.

27. J. J. Eggers, J. K. Su, and B. Girod, "Asymmetric watermarking schemes," *GI Jahrestagung Informatik 2000, Sicherheit in Mediendaten*, Berlin, Sept. 2000.

28. F.A.P. Petitcolas, "Watermarking schemes evaluation," *IEEE Signal Processing Magazine*, vol. 17, Sept. 2000.

29. F.A.P. Petitcolas and R.J. Anderson, "Evaluation of copyright marking systems," *IEEE Multimedia Systems*, Florence, Italy, vol. 1, pp. 574--579, June 1999.

30. F.A.P. Petitcolas, "unZign: is your watermark secure?," In http://www.cl.cam.ac.uk/~fapp2/watermarking/image_watermarking/unzign/

31. C. Candan, and N. Jayant, "A new interpretation of data hiding capacity," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing 2001*, Salt Lake City, May 2001 (in press).

32. "Theodore Roosevelt talks about Wilson and Taft" audio file, Vincent Voice Library, Michigan State U. Libraries, http://www.lib.msu.edu/vincent/t_roosevelt.ram

33. W.M. Fisher, G.R. Doddington, and K.M. Goudie-Marshall, "The DARPA speech recognition research database: Specifications and status," *Proc. DARPA Speech Recognition Workshop*, pp. 93-99, 1986.

34. C.G. Martin, "Digital image watermarking techniques," In **http://home.att.net/~steamedcrab/masterns.pdf**

35. Fco. J. Ruiz, A.A. Gokhale, and J.Y. Lee, Unpublished report, Class research project, ECE 966A, Michigan State U., East Lansing, Fall 1999.