Blind Watermark Estimation Attack for Spread Spectrum Watermarking

Hafiz Malik
Electrical and Computer Engineering Department
University of Michigan–Dearborn, Dearborn, MI 48128, USA
E-mail: *hafiz*@umich.edu, URL: http://www-personal.engin.umd.umich.edu/∼hafiz

Keywords: Spread-spectrum watermarking, independent component analysis, blind source separation, watermark estimation, detection, decoding

Received: September 18, 2008

This paper presents an efficient scheme for blind watermark estimation embedded using additive watermark embedding methods. The scheme exploits mutual independence between the host media and the embedded watermark and non-Gaussianity of the host media for watermark estimation. The proposed scheme employs the framework of independent component analysis (ICA) and poses the problem of watermark estimation as a blind source separation (BSS) problem. Analysis of the scheme shows that the proposed detector significantly outperforms existing correlation-based blind detectors traditionally used for SS-based watermarking. The proposed ICA-based blind detection/decoding scheme has been simulated using real-world audio clips. The simulation results show that the proposed ICA-based method can detect and decode watermark with extremely low decoding bit error probability (less than 0.01) against common watermarking attacks and benchmark degradations.

Povzetek: Opisana je metoda odkrivanja vodnega tiska.

1 Introduction

Digital forgeries and unauthorized sharing of digital media have emerged as a growing concern over the last decade. The widespread use of multimedia information is aided by factors such as the growth of the Internet, the proliferation of low-cost and reliable storage devices, the deployment of seamless broadband networks, the availability of state-of-the-art digital media production and editing technologies, and the development of efficient multimedia compression algorithms. Multimedia piracy has subjected the entertainment industry to enormous annual revenue losses. For example, music industry alone claims multi-million illegal music downloads on the Internet every week. It is therefore imperative to have robust technologies to protect copyrighted digital media from illegal sharing and tampering. Traditional digital data protection techniques, such as encryption and scrambling, alone cannot provide adequate protection as these technologies are unable to protect digital content once they are decrypted or unscrambled. Digital watermarking technology complements cryptography for protecting digital content even after it is deciphered [1].

Digital watermarking refers to the process of imperceptible embedding information (watermark) into the digital object (or the host object). Existing watermarking schemes based on the watermark embedding method used can be classified into two major categories:

1. *blind embedding*, in which the watermark embedder does not exploit the host signal information during watermark embedding process. Watermarking schemes based on spread-spectrum (SS) [1, 2, 3, 4, 5] fall into this category.

2. *informed embedding*, in which the watermark embedder exploits knowledge of the host signal during watermark embedding process. Watermarking schemes based on quantization index modulation [1, 6] belong to this category.

Similarity, existing watermarking schemes based on the detection method used cab be classified into two major categories:

1. *informed detector*, which assume that the host signal is available at the detector during watermark detection process, and

2. *blind detector*, which assume that the host signal is not available at the detector for watermark detection.

Although the performance expected from a given watermarking system depends on the target application area [1], but robustness of the embedded watermark and efficient detection are desirable features of a give watermarking scheme. In addition, fidelity (or imperception) of the embedded watermark is additional requirement of perception based watermarking schemes [1]. To meet fidelity requirement, the power of the embedded watermark (watermark strength) is generally kept much lower than the host signal power.

In this paper we consider additive watermark embedding model, e.g. SS-based watermarking, where the watermark signal is added to the host signal in the marking space to

obtain the watermarked signal. Existing watermark detection schemes for SS-based watermarking generally employ statistical characterization of the host signal to develop an optimal or suboptimal watermark detector [6, 7, 8]. It is important to mention that blind watermark detectors for SS-based watermarking perform poorly as the host-signal acts as interference at the blind decoder. Therefore, nonzero decoding error probability at the blind watermark decoder even in the absence of attack-channel distortion is one of the limitations of existing blind watermark detectors for SS-based watermarking schemes.

This paper presents a novel blind watermark detection method for the blind additive watermark embedding schemes[1, 2, 3, 4, 5]. The main motivation of this paper is to design a blind detector for SS-based watermarking schemes capable of suppressing host-signal interference (or improving watermark-to-host ratio) at the detector, hence improving decoding as well as detection performance. Towards this end, the proposed detector uses ICA framework by posing watermark detection problem as a blind source separation (BSS) problem. The proposed detector models the received watermarked signal as a linear mixture of underlying independent components (the host signal and the watermark). It also assumes non-Gaussianity of the host signal. Recently, we have shown in [15, 16, 17] that the watermark estimation problem for SS-based watermarking can be modeled as that of BSS of underdetermined mixture of independent sources. Therefore, the ICA framework could be used to estimate the watermark from the watermarked signals obtained using additive embedding model.

The proposed ICA-based detector first estimates the hidden independent components (i.e., the watermark and the host signal) from the received watermarked signal using the ICA framework, and then these estimated components are used to detect the embedded watermark. We present theoretical analysis to show that the proposed ICA-based detector performs significantly better than the existing watermark detectors operating without canceling the host signal interference at the watermark detector for watermark detection [6, 7]. Simulation results also show that the proposed detector in estimation-correlation based detection settings also outperforms the normalised correlation based detector (commonly used for watermark detection in SS-based watermarking community [1, 2, 3]) operating without host interference suppression. Simulation results presented in this paper are evaluated against variety of signal manipulations and degradations applied to the watermarked media. These signal degradations include addition of colored and white noise, resampling, requantization, lossy compression, filtering, time- and frequency-scaling, and StirMark for audio benchmark attacks [20, 19, 18]. The proposed ICA-based watermark detector is applicable to SS-based watermarking of all media types, i.e. audio, video and images. However, in this paper the proposed detector is tested for digital audio (which includes music and voiced speech signals only) as the host media for watermark embedding, detection, and performance analysis.

In the past ICA-based framework has been used for multimedia watermarking [9, 10, 11, 13, 14, 12]. However, existing ICA-based data-hiding schemes are either not applicable to SS-based watermarking [9, 10, 11, 13] or use an informed detection framework for watermark extraction/extraction [14, 12] therefore are not discussed in this manuscript. For example, Yu et al in [14] have proposed ICA-based watermark detector that can be used for SS-based watermarking but their detector uses the embedded watermark and a private data during watermark extraction process. Similarly, Sener et al's proposed ICA based watermark detector in [12] is also applicable to SS-based watermark detection, but their proposed detector also also requires the original watermark during watermark detection process; therefore, cannot be used for blind watermark detection/extraction applications.

Rest of the paper is organized as follow: basics of SS-based watermarking are discussed in Section 2; a brief overview of the independent component analysis theory is provided in Section 3. The proposed ICA-based watermark detector along with its decoding, detection, and maximum watermarking-rate performance analysis are described in Section 4. Simulation results for decoding bit error probability performance of the proposed ICA-based watermark detector and a correlation-based detector against different attacks and signal degradations are described in Section 5. Finally the concluding remarks along with future research directions are presented in Section 6.

2 Basics of SS-based watermarking

The SS based watermarking system can be modeled using a classical secure communication model [1], as shown in Fig. 1. In Fig. 1, $\mathbf{S} \in \mathcal{R}^n$ is a vector containing coefficients of the host signal in marking space. It is assumed that the coefficients, $S_i : i = 0, 1, \cdots, n-1$, are independent and identically distributed (i.i.d.) random variables (r.v.) with zero mean and variance σ_s^2. A watermark, $\mathbf{V}$, is generated using: (1) a message bit, $b \in \{\pm 1\}$, to be embedded into n coefficients of the host signal, (2) a key-dependent pseudo-random sequence $\mathbf{W} \in \{\pm 1\}^n$, and (3) a perceptual mask, $\alpha \in \mathcal{R}^n$, estimated based on the human auditory system (HAS) and the host signal $\mathbf{S}$, i.e. $\alpha = f(\mathbf{S}, \text{HAS})$. We further assume that the watermark sequence $\mathbf{W}$ and the host signal coefficients $\mathbf{S}$ are mutually independent. The amplitude-modulated watermark is spectrally shaped according to perceptual mask α to meet the fidelity requirement of the perception based watermarking. The watermarked signal $\mathbf{X}$ is obtained by adding an amplitude-modulated watermark $\mathbf{V} = \alpha \odot \mathbf{W}b$, here $\odot$ denotes element-wise product of the two vectors, to the host signal $\mathbf{S}$. The watermarked signal $\mathbf{X}$ can be expressed as

$$\mathbf{X} = \mathbf{S} + \mathbf{V}, \tag{1}$$

The embedding distortion, $\mathbf{D}_e$ can be expressed as,

$$\mathbf{D}_e = \mathbf{X} - \mathbf{S}. \tag{2}$$

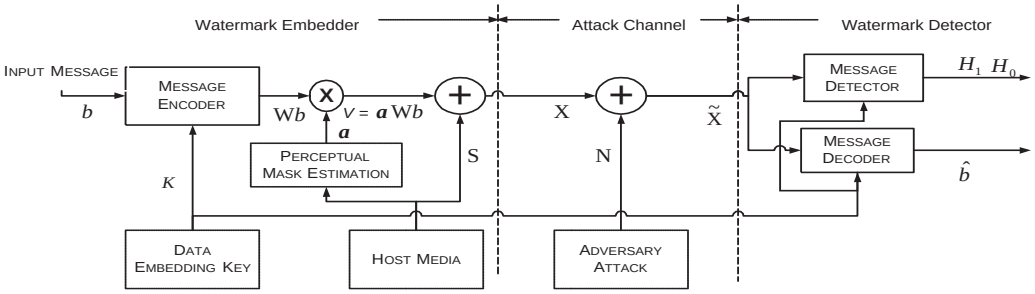

Figure 1: Perceptual based data hiding system with blind receiver

The mean-squared embedding distortion, d_e is expressed as,

$$\begin{aligned} d_e &= \frac{1}{n}E\{\|\mathbf{D}_e\|^2\} \\ &= \frac{1}{n}E\{\|\mathbf{X}-\mathbf{S}\|^2\} \\ &= \frac{1}{n}\|\alpha \odot \mathbf{W}b\|^2 \\ &= \frac{1}{n}\sum_{i=0}^{n-1}\alpha_i^2 = \sigma_v^2, \end{aligned} \tag{3}$$

where $\|\cdot\|$ represents the Euclidian norm, $E\{\cdot\}$ denotes expected value of a r.v., and σ_v^2 represents variance of the watermark $\mathbf{V}$.

The signal distortion due to an active adversary attack can be viewed as channel noise, $\mathbf{N}$, as shown in Fig. 1. The received watermarked signal at the detector, $\tilde{\mathbf{X}}$,

$$\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{N}, \tag{4}$$

is processed for watermark detection.

The watermarking schemes based on blind additive embedding model generally use probabilistic characterization of the host signal to develop an optimal or suboptimal watermark detector (in ML sense). The statistical characterizations of real-world host signal are available in spatial domain as well as in the transform domain. For example, stationary speech samples/coefficients both in the time domain and in the DWT domain can be approximated by Laplacian distribution [21] (see Appendix A for the probability distribution function (pdf) of DWT coefficients) i.e.,

$$f_s(\tau) = \frac{\beta}{2}e^{-\beta|\tau|}, \ \ |\tau| < \infty, \tag{5}$$

where $\beta = \frac{\sqrt{2}}{\sigma_s}$

The average decoding bit error probability, P_e, under zero-channel distortion scenario, i.e. $N_i = 0$, can be calculated by assuming that

1. the watermarked sample X_i is obtained by adding a binary amplitude-modulated watermark V_i, i.e. $\alpha_i W_i b$,
2. the detector is based on Neyman-Pearson criterion,
3. no pre-processing is applied to the watermarked audio to suppress host interference,
4. W_i takes values ± 1 with probability $\frac{1}{2}$,

In addition, for performance analysis we will consider two information embedding scenarios: (1) one bit $b \in \{\pm 1\}$ of information is embedded in each coefficient of the host signal, S_i, and (2) one bit $b \in \{\pm 1\}$ of information is embedding in $|\zeta|$ coefficients of the host signal $\mathbf{S}$, where $|\zeta|$ denotes the cardinality of the selected coefficient indices set ζ.

Consider one bit embedding per coefficient, i.e. $n = 1$, case first. It has been shown in [7] that the ML decoder estimates $\hat{b} = 1$ if $\tilde{X}_0 W_0 > 0$ and an error will occur when $\tilde{X}_0 W_0 < 0$. The average P_e is given by

$$\begin{aligned} P_e &= \Pr\{\tilde{X}_0 W_0 < 0 | b = 1\} \\ &= \int_{-\infty}^{0} f_s(\tau - \alpha)\, d\tau. \end{aligned} \tag{6}$$

Assuming the Laplacian distribution model for the host, it can be shown

$$P_e = \frac{1}{2}e^{-\sqrt{2}/\lambda_0}, \tag{7}$$

where $\lambda_0 = \frac{\sigma_s}{\sigma_{v_0}}$ which is generally referred as *signal-to-watermark ratio* (SWR), when expressed in dB i.e. $SWR = 20\log_{10}\lambda$.

It can be observed from Eq. (7) that non-zero P_e is not achievable even in the absence of attack-channel distortion, and $P_e = f(\lambda)$. In addition, the value of the parameter λ determines the tradeoff between fidelity of the embedded watermark and P_e.

Consider second embedding scenario, i.e., one bit information is embedded in $|\zeta| = n$ coefficients of the host. In this case the watermarked audio is given by,

$$X_i = S_i + \alpha_i W_i b, \ i \in \zeta. \tag{8}$$

Let us assume that the watermarked signal used for detection is free of attack-channel distortion, and message symbols are equally probable. In this case, the ML decoder that minimizes the decoding error probability will assign decision regions D_- and D_+ as follow,

$$\ln\frac{f_x(\mathbf{x}|b_+)}{f_x(\mathbf{x}|b_-)} = \ln\frac{f_x(\mathbf{x}-\hat{\alpha}\mathbf{w})}{f_x(\mathbf{x}+\hat{\alpha}\mathbf{w})} \underset{D_-}{\overset{D_+}{\gtrless}} 0, \tag{9}$$

where $b_+(resp.b_-)$ represent the event that binary information $b = +1(resp.b = -1)$ is embedded in the selected indices and $\hat{\alpha}$ is the masking threshold estimated from watermarked audio. It is shown in Section 4 that the estimated of masking threshold from the unwatermarked and watermarked audio clip are very close given that attack-channel distortion induced into the watermarked audio is below certain threshold. It is therefore reasonable to assume that $\hat{\alpha} \approx \alpha$.

The ML sufficient statistic, T, assuming Laplacian pdf for the host coefficients S_i, can be written as,

$$T(\mathbf{x}|\ \mathbf{s},\ \hat{\alpha}) = \sum_{i\in\zeta} \beta\left(|X_i + \hat{\alpha}_i W_i| - |X_i - \hat{\alpha}_i W_i|\right). \quad (10)$$

If $b = 1$ was embedded, then the sufficient statistics T can be expressed as,

$$T(\mathbf{x}|\ \mathbf{s},\ \hat{\alpha}) = \sum_{i\in\zeta} \beta\left(|S_i + 2\hat{\alpha}_i W_i| - |S_i|\right). \quad (11)$$

Here the ML detector is a bit-by-bit hard decoder, i.e.,

$$\hat{b} = \text{sgn}(T). \quad (12)$$

To determine the bit error probability for this ML decoder, a statistical characterization of T is required. Here T is sum of $|\zeta|$ i.i.d. random variables. Therefore, by applying the central limit theorem (CLT), T can be approximated by the Gaussian random variable. Mean of T, $E\{T\}$ can be calculated as,

$$E\{T(\mathbf{x}|\ \mathbf{s},\ \hat{\alpha})\} = \sum_{i\in\zeta} \beta\left(E_{s,w}\left(|S_i + 2\hat{\alpha}_i W_i| - |S_i|\right)\right), \quad (13)$$

and variance,

$$E\{T\} = \sum_{i\in\zeta} \left(e^{-2\sqrt{2}/\lambda_i} + \frac{2\sqrt{2}}{\lambda_i} - 1\right), \quad (14)$$

$$\text{Var}\{T(\mathbf{x}|\ \mathbf{s},\ \hat{\alpha})\} = \sum_{i\in\zeta} \beta\left(\text{Var}_{s,w}\left(|S_i + 2\hat{\alpha}_i W_i| - |S_i|\right)\right) \quad (15)$$

$$\text{Var}\{T\} = \sum_{i\in\zeta} \left(3 - e^{-4\sqrt{2}/\lambda_i} - e^{-2\sqrt{2}/\lambda_i}\left(1 + \frac{4\sqrt{2}}{\lambda_i}\right)\right). \quad (16)$$

In this case, the P_e is given as,

$$P_e = Q\left(\frac{|E\{T\}|}{\sqrt{\text{Var}\{T\}}}\right), \quad (17)$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt$

Eq. (17) shows that the decoding error probability P_e is non-zero even in the absence of attack-channel distortion, and P_e is a function of λ. The above analysis also shows that the detection/decoding performance of a blind detector for additive embedding schemes is inherently bounded below by the host-signal interference at the detector. The main motivation behind this paper is to design a watermark detector for additive embedding schemes with an improved watermark detection, decoding, and maximum watermarking-rate performances by suppressing the host-signal interference at the blind detector. Towards this end, theory of ICA is used by posing watermark estimation for additive embedding as a BSS problem. The proposed framework first estimates embedding watermark using BSS based on ICA which is then used for detection and decoding. The fundamentals of the ICA theory are briefly outlined in the following section followed by the details of the proposed ICA-based detector.

3 Independent Component Analysis

Independent component analysis (ICA) is a statistical framework for estimating underlying hidden factors or components of multivariate statistical data. In the ICA model, the data variables are assumed to be linear or nonlinear mixtures of some unknown latent variables, and the mixing system is also unknown [23, 22]. The hidden variables are also assumed to be non-Gaussian and mutually independent. The ICA model can be considered as an extension of the principal component analysis (PCA) and factor analysis [23, 22]. In fact, ICA can be treated as non-Gaussian factor analysis, since data is modeled as a linear mixture of underlying non-Gaussian factors. The ICA framework has been used in diverse application scenarios including blind source separation (BSS), feature extraction, telecommunication, and economics [23, 22]. In the following we will review only the linear ICA framework since only that is relevant to the SS-based watermarking model. In general, the linear ICA model can be defined for noise-free as well as noisy scenarios as follows.

Noise-free ICA model: *ICA of a random vector* $\mathbf{X} \in \mathcal{R}^m$ *consists of estimating the following generative model of the data:*

$$\mathbf{X} = \mathbf{AS}, \quad (18)$$

where $\mathbf{X}$ *represents* n*-realizations of the observed* m*-dimensional random vector,* $\mathbf{S} \in \mathcal{R}^{n_1}$ *is the hidden random variables and* $\mathbf{A} \in \mathcal{R}^{m\times n_1}$ *is mixing matrix. The hidden variables,* $\mathbf{S}^{(i)}$*, in the vector* $\mathbf{S} = \left[\mathbf{S}^{(1)}, \cdots, \mathbf{S}^{(n_1)}\right]^{\mathrm{t}}$ *are assumed statistically independent.*

Noisy ICA model: *ICA of a random vector* $\mathbf{X}$ *consists of estimating the following generative model of the data:*

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}, \quad (19)$$

where $\mathbf{N}$ *is* n*-realizations of an* m*-dimensional random noise, while* $\mathbf{X}$, $\mathbf{S}$, and $\mathbf{A}$ *are the same as in the noise-free model in* Eq. (18).

In this paper, we use the noisy ICA generative model to design an ICA-based watermark detector for SS-based watermarking schemes. The proposed ICA-based watermark detector attempts to estimate the embedded watermark from the watermarked signal while reducing the host-signal interference at the watermark detector. Before estimating the underlying independent components from observed data using ICA framework, the generative model should meet certain conditions to ensure the identifiability of the ICA model. The identifiability constraints defined in [22, 24, 25, 29, 26, 27] underdetermined ICA (UICA) model are outlined below:

1. *Statistical independence*: The hidden (latent) variables/sources are statistically independent.
2. *Non-Gaussianity*: At most one of the underlying independent components $\mathbf{S}^{(i)},\ i = 1, 2 \cdots n_1$, is normally distributed.

Therefore, independence and maximum non-Gaussianity are two fundamental ingredients of the UICA framework. Independence of the underlying components is one of the assumptions that is made to estimate components from the linear mixture. Note that independence of the underlying components is a stronger condition than uncorrelatedness, e.g., for the BSS problem, there might be many dependent but uncorrelated representations of the observed signals and these uncorrelated but dependent representations of the observed signals cannot separate the mixed sources [22]. Therefore, uncorrelatedness itself is insufficient to solve the BSS problem. In fact, independence implies nonlinear uncorrelatedness [22], that is, if $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are two independent components then any nonlinear transformations of these components, say, $\phi_1(\mathbf{S}^{(1)})$ and $\phi_2(\mathbf{S}^{(2)})$, are uncorrelated as well (i.e. their covariance is zero). On the other hand, if $\mathbf{S}^{(1)}$ and $\mathbf{S}^{(2)}$ are assumed to be just uncorrelated then in general, the corresponding nonlinear transformations do not necessarily have zero covariance. Thus to perform ICA, a stronger form of decorrelation of the underlying components is required, that is, nonlinear decorrelation. A suitable selection of nonlinearities, i.e. $\phi_1(\cdot)$ and $\phi_2(\cdot)$, can be achieved by using tools like maximum likelihood and mutual information from estimation theory and information theory [22].

Maximum non-Gaussianity is another important requirement of ICA-based hidden components estimation [23, 22, 38, 30]. A quantity kurtosis defined in terms of the fourth-order central moment κ is generally used as a measure of non-Gaussianity of a random variable. Kurtosis of a real random variable S can be defined as,

$$\kappa = \left(E\{(S - E(S))^4\} / E^2\{(S - E(S))^2\} \right) - 3. \quad (20)$$

A normal random variable has zero kurtosis; therefore, kurtosis is a measure of the *distance* of a random variable from a Gaussian distribution. Distributions that are peakier (flatter) about the mean than a Gaussian distribution generally have positive (negative) kurtosis. Random variables with positive kurtosis, i.e. $\kappa > 0$, are generally called super-Gaussian. The Laplacian distribution is a typical example of this case. Random variables with negative kurtosis value, i.e., $\kappa < 0$ are called sub-Gaussian, e.g., the uniform distribution.

The BSS is one of the most widely explored applications of the ICA model [23, 22]. In case of BSS using ICA framework, the recovery of the underlying sources relies on the assumption that the constituent sources are mutually independent. The *cocktail party problem* is a classical example of BSS, where several people are simultaneously speaking in the same room and objective is to separate voices of different speakers using microphone recordings (in the room). In order to illustrate the idea n_1 speakers (sources) are considered here. The observation $\mathbf{X} \in \mathcal{R}^{m \times\ n}$ is generated by mixing sources $\mathbf{S} \in \mathcal{R}^{n_1 \times\ n}$ by a *mixing matrix* $\mathbf{A} \in \mathcal{R}^{m \times\ n_1}$. The static linear mixing model can be expresses as,

$$\mathbf{X}_i = \mathbf{A}\mathbf{S}_i + \mathbf{N}_i,\ i = 1, 2, \cdots,\ n \quad (21)$$

The aim of BSS is the recover the underlying sources $\mathbf{S}^{(l)},\ l = 1, 2, \cdots, n_1$ from the observation $\mathbf{X}$ only. The ICA achieves the separation relying on the assumption that the underlying sources are mutually independent. To this end the ICA framework finds a linear representation in which the underlying components are statistically independent. In other words, BSS using ICA tries to estimate the *demixing* (*separating*) matrix, $\mathbf{B} \in \mathcal{R}^{n_1 \times m}$, from the observed data $\mathbf{X}$. The estimated demixing matrix is the inverse (or generalized inverse) of mixing matrix $\mathbf{A}$, i.e., $\hat{\mathbf{B}} = \hat{\mathbf{A}}^{\dagger} = (\hat{\mathbf{A}}^T \hat{\mathbf{A}})^{-1} \hat{\mathbf{A}}^T$. Most of existing BSS schemes using ICA model are based on the information-theoretic framework. For example, Bell et al's [21] ICA scheme is based on the idea of information maximization, or *infomax* among the estimated independent components. P. Comon in [23] has used higher-order cumulants whereas, Gaeta et al in [28] used ML estimation framework for BSS. Many existing BSS methods are extensions of infomax, higher-order cumulants, and ML method [23, 22].

4 Proposed ICA Based Watermark Detector

The proposed ICA-based watermark detector consists of two stages: 1) watermark estimation stage, and, 2) watermark decoding and/or detection stage. The watermark estimation stage estimates watermark $\tilde{\mathbf{V}}$ from the received watermarked audio $\tilde{\mathbf{X}}$ using ICA framework, whereas, the watermark decoding (resp. detection) stage decodes (resp. detects) the embedded watermark using the ML approach. The block diagram of the proposed watermark detector is given in the Fig. 2.

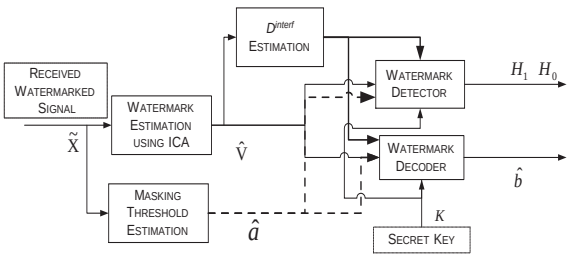

Figure 2: Block diagram of the proposed ICA-based watermark detector

In general the ICA model for BSS estimates the demixing matrix $\hat{\mathbf{B}}$ from the observed data $\mathbf{X}$, and hence the underlying independent components $\hat{\mathbf{S}}^{(i)}$. This model is ex-

tendable to the watermark estimation problem for the watermarked signal, assuming identifiability conditions of the UICA model are satisfied. To verify whether the additive embedding model (Eq. (1)) satisfies the identifiability constraints of an ICA model, rewrite Eq. (1) with $b = 1$, i.e.,

$$\mathbf{X} = \mathbf{S} + \alpha \odot \mathbf{W}.$$

The non-Gaussianity of the host signal and the watermark is the only requirement to satisfy constraints of UICA. As mentioned in Section (2) that real-world audio samples/coefficients in the time domain as well as in the DWT domain can be approximated by the Laplacian distribution (see Appendix A) and therefore if the watermark, $\mathbf{W}$, is generated based on some non-Gaussian distribution then non-Gaussianity constraint of UICA model is satisfied as well. Once the identifiability conditions of the UICA model are satisfied, the noisy ICA model can be extended to estimate the watermark from the watermarked audio generated using Eq. (1).

4.1 Watermark Estimation

For watermark estimation, the proposed watermark detector first estimates the watermark-mixing matrix $\hat{\mathbf{A}}$ which is then to estimate the underlying independent components (i.e., the host signal $\mathbf{S}$ and the watermark $\mathbf{W}$). An estimate of the watermark-mixing matrix, $\hat{\mathbf{A}}$, is usually obtained by optimizing a highly nonlinear function of the hidden sources also known as contrast function [23, 22]. The pseudo-inverse of the estimated watermark-mixing matrix $\hat{\mathbf{A}}^{\dagger}$ is applied to the observed mixture to estimate the host signal $\hat{\mathbf{S}}$ and the watermark $\hat{\mathbf{W}}$. However, as noted earlier, in the case of blind detectors for SS-based watermarking schemes, watermark estimation using ICA framework is a degenerate case, i.e., $m < n_1$. Therefore, just the estimation of watermark-mixing matrix is insufficient to separate the underlying independent components perfectly. In the case of additive embedding, the equation $\mathbf{X} = \hat{\mathbf{A}}\mathbf{S}$ has an affine set of solutions [34]. A preferred solution in this affine set is generally selected using probabilistic prior model of the independent components [39]. The performance of the proposed ICA-based watermark estimator depends on the separation quality of the separated (estimated) watermark. The separation quality of the separated source is generally measured in terms of, 1) *source-to-interference ratio* (*watermark-to-interference ratio (WIR)*, in case of watermark estimation), *source-to-noise ratio*, and 2) *source-to-artifact ratio* (for further details on these separation quality measures please see [35] and references therein). For performance analysis of the proposed ICA detector, only WIR distortion measure is considered here; therefore, the estimated watermark can be expressed as

$$\hat{V}_i = \eta_{1i}\alpha_i W_i b + S_i^{\text{interf}}, \tag{22}$$

where $\eta_{1i} \in \mathcal{R},\ 0 < \eta_{1i} \leq 1$ and S_i^{interf} is interference due to the host signal.

Let $S_i^{\text{interf}} = \eta_{2i} S_i,\ \eta_{2i} \in \mathcal{R}$, and $0 < \eta_{2i} \leq 1$ then Eq. (22) can be rewritten as,

$$\hat{V}_i = \eta_{1i}\alpha_i W_i b + \eta_{2i} S_i. \tag{23}$$

The relative distortion due to interference in the estimated watermark is defined as,

$$D^{\text{interf}} = \left(\eta_1/\eta_2\right)^2, \tag{24}$$

where $WIR = 10\log_{10}\left(D^{\text{interf}}\right)$ dB.

In general, $D^{\text{interf}} > 0$ dB for most of existing BSS schemes based on ICA framework [34, 35]. Several researchers have proposed elegant BSS algorithms based on ICA model for noisy data [38, 36, 31], these algorithms can be used for watermark estimation from the watermark audio. Among these, the FastICA for noisy data [38] is used in this paper due to its better computational and separation quality performance over existing algorithms [34].

It can be observed from Eq. (23) that the ICA stage acts as a pre-processing stage that suppresses the host interference or improves *watermark-to-host ratio*. Once estimated watermark $\hat{\mathbf{V}}$ is available, an optimal detector can be designed based on the statistics of $\hat{\mathbf{V}}$ for watermark detection (resp. decoding). It is important to notice that ICA based pre-processing stage uses constraints like mutual independence of the underlying sources, non-Gaussianity, and multichannel observation i.e. $m \geq 2$. A constrained optimization of highly nonlinear cost function e.g. $\tanh(x)$, $x\exp(-x^2)$, etc. is used to suppress the host interference in the estimated watermark [22, 23]. In addition, under practical scenarios, BSS using ICA also requires reasonably large number of data samples n to separate the underlying sources. Therefore, ICA based pre-processing to suppress host interference is inherently different from filtering based pre-processing schemes i.e., optimal linear filtering [44], wiener filtering, non-linear filtering, etc. The ICA-based pre-processing stage is to improve *watermark-to-host ratio* hence expected to improve the detection performance [41, 42]. It is however important to mention that improvement comes at the cost of higher computational power.

In the following subsections we analyze the performance of the proposed ICA-based detector in terms of three parameters: (1) detection rate in terms of false positives and true positives 4.2, (2) decoding error probability 4.3, and (3) maximum watermarking rate 4.4.

4.2 Watermark Detection: *Performance Analysis*

A watermark detector is generally characterized by two performance measures: the probability of *false alarm* P_F and the probability of *detection* P_D. The probability of detection represents the probability of deciding on the presence of a watermark when the received audio indeed contains a watermark. The probability of false alarm represents chances of deciding the presence of a watermark when in

fact the received audio does not contain a watermark. The watermark detection process can be treated as a binary decision problem in the presence or absence of attack-channel distortions.

We first consider the case where the received watermarked audio has not suffered attack-channel distortion. The estimated watermark is given by,

$$\hat{V}_i = \eta_{1i}\alpha_i W_i b + \eta_{2i} S_i,\ i \in \zeta. \tag{25}$$

In this scenario, the watermark detection can be formulated as a binary hypotheses test,

$$\begin{aligned} H_1 : \hat{V}_i &= \eta_{1i}\hat{\alpha}_i W_i b + \eta_{2i} S_i \\ H_0 : \hat{V}_i &= \eta_{2i} S_i,\ i \in \zeta. \end{aligned} \tag{26}$$

In this detection problem, the watermark $\mathbf{W}$ is the target signal and host interference, $\eta_2\mathbf{S}$ acts as additive noise. The goal of watermark detector is to determine presence or absence of the watermark in the estimated watermark $\hat{\mathbf{V}}$ based on the statistics of $\mathbf{S}$ and $\mathbf{W}$. Let us assume that statistics of unwatermarked and watermarked audio are same [43], therefore *pdfs* under each hypothesis are known. The decision rule, in this scenario is based on likelihood ratio which is given as:

$$\Lambda(\hat{\mathbf{V}}) = \prod_{\zeta} \left(\frac{f_{\hat{v}}(\hat{\mathbf{v}}|H_1)}{f_{\hat{v}}(\hat{\mathbf{v}}|H_0)} \right) \underset{H_0}{\overset{H_1}{\gtrless}} \acute{\xi} \tag{27}$$

where $\Lambda(\hat{\mathbf{V}})$ is likelihood ratio and $\acute{\xi}$ is decision threshold.

The log-likelihood is defined as,

$$\begin{aligned} L(\hat{\mathbf{V}}) &= \ln\left(\Lambda(\hat{\mathbf{V}})\right) \qquad (28) \\ &= \ln\left(\prod_{\zeta} \left(\frac{f_{\hat{v}}(\hat{\mathbf{v}}|H_1)}{f_{\hat{v}}(\hat{\mathbf{v}}|H_0)} \right) \right) \\ &= \ln\left(\prod_{\zeta} \left(\sum_{l} \left(\frac{p(b_l) f_{\hat{v}}(\hat{\mathbf{v}}|H_1)}{f_{\hat{v}}(\hat{\mathbf{v}}|H_0)} \right) \right) \right) \\ &= \ln\left(\prod_{\zeta} \left(\sum_{l} \left(\frac{p(b_l) f_{\acute{s}}(\hat{\mathbf{v}} - \hat{\alpha} \odot \mathbf{w} b_l)}{f_{\acute{s}}(\hat{\mathbf{v}})} \right) \right) \right) \\ &\underset{H_0}{\overset{H_1}{\gtrless}} \xi \qquad (29) \end{aligned}$$

where, $l \in \{\pm 1\}$, $\xi = \ln(\acute{\xi})$ and r.v. $\acute{S}_i$ is defined as $\acute{S}_i = \eta_{2i} S_i$.

In the above test, the decision threshold ξ can be minimized based on Neyman-Pearson rule, that is, maximize the P_D for a given value of P_F [41, 42].

Assuming Laplacian distribution for the host audio, the $L(\hat{\mathbf{V}})$ can be written as,

$$L(\hat{\mathbf{V}}|\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2) = \sum_{i\in\zeta} \acute{\beta}_i \left(|\hat{V}_i| - |\hat{V}_i - \hat{\eta}_{1i}\hat{\alpha}_i W_i| \right), \tag{30}$$

where $\acute{\beta}_i = \sqrt{2}/\hat{\eta}_{2i}\sigma_s$, $\hat{\eta}_1$, and $\hat{\eta}_2$ are estimates of scaling coefficients of $\mathbf{V}$ and $\mathbf{S}$ in $\hat{\mathbf{V}}$. Estimation details of $\hat{\eta}_1$, and $\hat{\eta}_2$ are discussed in Section 4.5.

The statistical characterization of $L(\hat{V})$ under hypothesis H_0 can be determined as,

$$L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2) = \sum_{i\in\zeta} \acute{\beta}_i \left(|\hat{V}_i| - |\hat{V}_i - \hat{\eta}_{1i}\hat{\alpha}_i W_i| \right), \tag{31}$$

$$L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2) \overset{\text{def}}{=} \sum_{i\in\zeta} \acute{\beta}_i\left(Z_i\right), \tag{32}$$

$$\text{where } Z_i \overset{\text{def}}{=} |\hat{V}_i| - |\hat{V}_i - \hat{\eta}_{1i}\hat{\alpha}_i W_i|. \tag{33}$$

Here $L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)$ is the sum of $|\zeta|$ statistically independent random variables that can be approximated by the Gaussian random variable based on the CLT, mean, m_0 and variance, σ_0^2 of $L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)$ is calculated as follows,

$$\begin{aligned} m_0 &\overset{\text{def}}{=} E\{L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} \qquad (34) \\ &= \sum_{i\in\zeta} \acute{\beta}_i E\{Z_i\}, \\ \sigma_0^2 &\overset{\text{def}}{=} \text{Var}\{L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} \qquad (35) \\ &= \sum_{i\in\zeta} \acute{\beta}_i^2 \text{Var}\{Z_i\}. \end{aligned}$$

Averaging Eq. (36) over W, we have,

$$E_w\{Z_i\} = |\acute{S}_i| - \frac{1}{2}\left(|\acute{S}_i| + \hat{\eta}_{1i}\hat{\alpha}_i + \left| |\acute{S}_i| - \hat{\eta}_{1i}\hat{\alpha}_i \right| \right), \tag{36}$$

$$\text{Var}_w\{Z_i\} = \frac{1}{4}\left(|\acute{S}_i| + \hat{\eta}_{1i}\hat{\alpha}_i + \left| |\acute{S}_i| - \hat{\eta}_{1i}\hat{\alpha}_i \right| \right)^2. \tag{37}$$

These equation can be rewritten as,

$$E_w\{Z_i\} = \begin{cases} |\acute{S}_i| - \hat{\eta}_{1i}\hat{\alpha}_i & |\acute{S}_i| \le \hat{\eta}_{1i}\hat{\alpha}_i \\ 0 & |\acute{S}_i| > \hat{\eta}_{1i}\hat{\alpha}_i, \end{cases} \tag{38}$$

$$\text{Var}_w\{Z_i\} = \begin{cases} |\acute{S}_i^2| & |\acute{S}_i| \le \hat{\eta}_{1i}\hat{\alpha}_i \\ \hat{\eta}_{1i}^2\hat{\alpha}_i^2 & |\acute{S}_i| > \hat{\eta}_{1i}\hat{\alpha}_i. \end{cases} \tag{39}$$

Averaging it over $\acute{S}$ we have,

$$\begin{aligned} E\{Z_i\} &= E_{\acute{s}}\left(E_w\{Z_i\}\right) \qquad (40) \\ &= \frac{1}{\acute{\beta}_i}\left(1 - e^{-\acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i} - \acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i\right), \end{aligned}$$

$$\begin{aligned} &\text{Var}\{Z_i\} = E_{\acute{s}}\left(\text{Var}_w\{Z_i\}\right) + \text{Var}_{\acute{s}}\left(E_w\{Z_i\}\right) \qquad (41) \\ &= \frac{1}{\acute{\beta}_i^2}\left(3 - e^{-2\acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i} - 2e^{-\acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i}\left(1 + 2\acute{\beta}_i\hat{\eta}_i\hat{\alpha}_i\right)\right). \end{aligned}$$

Substituting $E\{Z_i\}$, and $\text{Var}\{Z_i\}$ in Eq. (36), we have,

$$m_0 = \sum_{i\in\zeta}\left(1 - e^{-\frac{\hat{\eta}_{1i}\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}} - \frac{\hat{\eta}_{1i}\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}\right), \tag{42}$$

$$\sigma_0^2 = \sum_{i\in\zeta}\left(3 - e^{-\frac{\hat{\eta}_{1i}2\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}} - 2e^{-\frac{\hat{\eta}_{1i}\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}}\left(1 + \frac{2\hat{\eta}_{1i}\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}\right)\right). \tag{43}$$

Similarly, $L(\hat{V})$ under hypothesis H_i can be written as,

$$L(\hat{\mathbf{V}}|H_1,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2) = \sum_{i\in\zeta} \acute{\beta}_i \left(|\hat{V}_i + \hat{\eta}_{1i}\hat{\alpha}_i W_i| - |\hat{V}_i| \right) \tag{44}$$

Here $L(\hat{\mathbf{V}})$ can be approximated by a Gaussian random variable with the same set of assumptions as under hypothesis H_0. In addition, the distribution of $L(\hat{\mathbf{V}})$ under hypothesis H_1 is symmetrical to the distribution of under H_0 with respect to the origin. Therefore,

$$\begin{aligned} m_1 &\stackrel{\text{def}}{=} E\{L(\hat{\mathbf{V}}|H_1,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} \quad (45)\\ &= -E\{L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\},\\ \sigma_1^2 &\stackrel{\text{def}}{=} \text{Var}\{L(\hat{\mathbf{V}}|H_1,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} \quad (46)\\ &= \text{Var}\{L(\hat{\mathbf{V}}|H_0,\ \mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} \end{aligned}$$

Now the probability of false alarm P_F and the probability detection P_D are given as,

$$P_F = Q\left(\frac{\xi + m_1}{\sigma_1}\right), \quad (47)$$

$$P_D = Q\left(\frac{\xi - m_1}{\sigma_1}\right). \quad (48)$$

Lets define the watermark-to-noise ratio (*WNR1*) as

$$WNR1 \stackrel{\text{def}}{=} \frac{m_1^2}{\sigma_1^2}. \quad (49)$$

If we denote by $Q^{-1}(P_F)$ the value $x \in \mathcal{R}$ such that $Q(x) = P_F$ then *receiver operating characteristics (ROC)* of the proposed detector can be expressed as:

$$P_D = Q\left(Q^{-1}(P_F) - 2\sqrt{WNR1}\right). \quad (50)$$

It can be observed from Eq. (50) that the detection performance of the proposed detector is a function of WNR1. Since the proposed ICA-based detector is designed to reduce the host-signal interference before detection, therefore, the ICA-based detector is expected to perform better than the existing blind detectors [6, 3, 1] operating without reducing host signal interference. The detection performance improvement can be attributed to its host interference suppression or watermark-to-host ratio improving capability. To illustrate this notion the theoretical ROC performance of the proposed detector based on Eq. (50) for different values of host interference suppression values (or *WIR*) is given in Fig. 3. It can be observed from Fig. 3 that the proposed detector performs superior that the detector operating without host interference canceling.

4.3 Watermark Decoding: *Performance Analysis*

To evaluate performance of the proposed detector in terms of decoding error probability, let us consider watermark embedding model given in Eq. (1) and decoding framework discussed in Section 2. Consider one bit per coefficient embedding case first, that is, $X_0 = S_0 + \alpha_0 W_0$. The P_e in this case for $\hat{V}_0$ can be expressed as,

$$P_{e\_ICA} = \frac{1}{2} e^{-\{\frac{\hat{\eta}_{10}}{\hat{\eta}_{20}}\}\left(\sqrt{2}/\lambda_0\right)}. \quad (51)$$

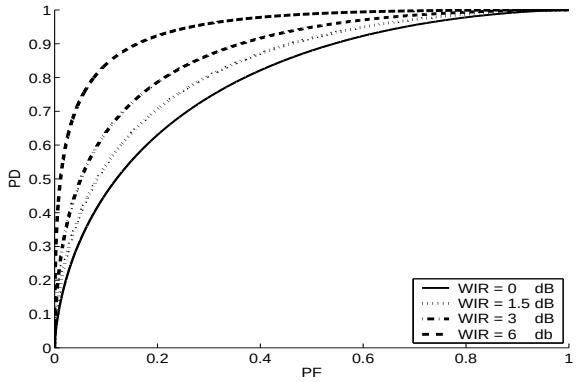

Figure 3: ROC performance of the proposed detector for different values of *WIR*, SWR = 13 dB, and one bit per $|\zeta|$ coefficients embedding, where $|\zeta| = 5$ (theoretical values)

Here Eq. (51) shows that ICA-based detector does improve decoding error performance. The decoding error performance gain for the proposed ICA-based detector over the traditional detector can expressed,

$$G = \frac{P_e}{P_{e\_ICA}} = e^{-\frac{\sqrt{2}}{\lambda_0}\{1 - \frac{\hat{\eta}_{10}}{\hat{\eta}_{20}}\}}. \quad (52)$$

It is important to mention that in general BSS using ICA have relatively small interference distortion, i.e., $\hat{\eta}_{10}/\hat{\eta}_{20}$ [34], therefore, $G \geq 0$ for $WIR > 0$dB. The performance gain ot the proposed ICA-based detector over that of the decoder given by Eq. (7) is plotted in Fig. 4. It can be observed from Fig. 4 that for a fixed value of *SWR*, decoding error probability of the proposed detector improves with the increase in the separation quality of the ICA scheme used for watermark estimation.

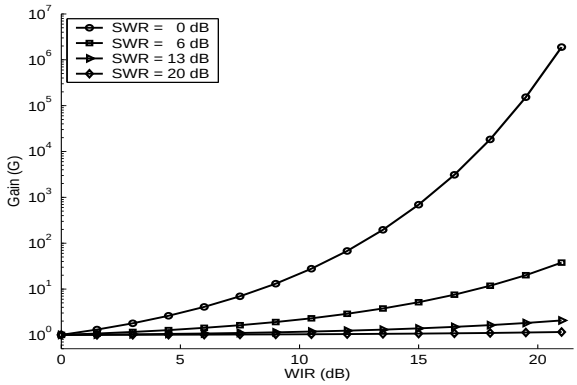

Figure 4: The decoding performance gain due to host-interference suppression at the detector (theoretical values)

Now consider second embedding scenario, that is, one bit is embedded into $|\zeta|$ coefficients of the host signal $\mathbf{S}$, i.e.,

$$X_i = S_i + \hat{\alpha}_i W_i b,\ i \in \zeta. \quad (53)$$

In this case, the estimated watermark $\hat{\mathbf{V}}$ using proposed ICA-based watermark detector, under zero attack-channel

distortion, can be expressed as,

$$\hat{V}_i = \eta_{2i} S_i + \eta_{1i}\hat{\alpha}_i W_i b,\ i \in \zeta. \tag{54}$$

For equally probable message symbols the ML decoder that minimizes the P_e will satisfy the following condition,

$$\ln \frac{f_{\hat{v}}(\hat{\mathbf{v}}|b_+, \hat{\alpha}, \hat{\eta}_1, \hat{\eta}_2)}{f_{\hat{v}}(\hat{\mathbf{v}}|b_-, \hat{\alpha}, \hat{\eta}_1, \hat{\eta}_2)} = \ln \frac{f_{\acute{s}}(\hat{\mathbf{v}} - \hat{\eta}_1\hat{\alpha}\mathbf{w})}{f_{\acute{s}}(\hat{\mathbf{v}} + \hat{\eta}_1\hat{\alpha}\mathbf{w})} > 0. \tag{55}$$

The ML sufficient statistics for Laplacian $\acute{\mathbf{S}}$ can be written as,

$$T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2) = \tag{56}$$
$$\sum_{i\in\zeta} \acute{\beta}_i \left(|\hat{V}_i + \hat{\eta}_{1i}\hat{\alpha}_i W_i| - |\hat{V}_i - \hat{\eta}_{1i}\hat{\alpha}_i W_i| \right)$$

Assuming $b = 1$, then $T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)$ can be expressed as,

$$T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2) = \sum_{i\in\zeta} \acute{\beta}_i \left(|\acute{S}_i + 2\hat{\eta}_{1i}\hat{\alpha}_i W_i| - |\acute{S}_i| \right), \tag{57}$$

$$T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2) \ \stackrel{\text{def}}{=}\ \sum_{i\in\zeta} \acute{\beta}_i Z_i, \tag{58}$$

where, $Z_i \stackrel{\text{def}}{=} \left(|\acute{S}_i + 2\hat{\eta}_{1i}\hat{\alpha}_i W_i| - |\acute{S}_i| \right)$ The ML decoder is a bit-by-bit hard decoder

$$\hat{b} = \text{sgn}(T) \tag{59}$$

To determine the P_e for the ML decoder, a statistical characterization of $T(\hat{\mathbf{V}})$ is required. As $T(\hat{\mathbf{V}})$ is sum of $|\zeta|$ i.i.d. random variables, therefore, using CLT, $T(\hat{\mathbf{V}})$ can be approximated by the Gaussian random variable, the mean and variance of T can be computed as,

$$E\{T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} \ \stackrel{\text{def}}{=}\ \sum_{i\in\zeta} \acute{\beta}_i E\{Z_i\}, \tag{60}$$

$$\text{Var}\{T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} \ \stackrel{\text{def}}{=}\ \sum_{i\in\zeta} \acute{\beta}^2 \text{Var}\{Z_i\}. \tag{61}$$

In Z_i, W and $\acute{S}$ are the only r.v.s, so averaging Z_i over r.v. W condition to the selected host indices $\acute{S}$ and $W_i \in \{\pm 1\}$ with probability $\frac{1}{2}$ we have,

$$E_w\{Z_i\} = \frac{1}{2}\left(|\acute{S}_i| + 2\hat{\eta}_{1i}\hat{\alpha}_i + \left||\acute{S}_i| - 2\hat{\eta}_{1i}\hat{\alpha}_i\right| \right) - |\acute{S}_i|, \tag{62}$$

$$\text{Var}_w\{Z_i\} = \frac{1}{4}\left(|\acute{S}_i| + 2\hat{\eta}_{1i}\hat{\alpha}_i + \left||\acute{S}_i| - 2\hat{\eta}_{1i}\hat{\alpha}_i\right| \right)^2. \tag{63}$$

rewriting the above equations, we have,

$$E_w\{Z_i\} \ = \ \begin{cases} -|\acute{S}_i| + 2\hat{\eta}_{1i}\hat{\alpha}_i & |\acute{S}_i| \le 2\hat{\eta}_{1i}\hat{\alpha}_i \\ 0 & |\acute{S}_i| > 2\hat{\eta}_{1i}\hat{\alpha}_i, \end{cases} \tag{64}$$

$$\text{Var}_w\{Z_i\} \ = \ \begin{cases} |\acute{S}_i^2| & |\acute{S}_i| \le 2\hat{\eta}_{1i}\hat{\alpha}_i \\ 4\hat{\eta}_{1i}^2\hat{\alpha}_i^2 & |\acute{S}_i| > 2\hat{\eta}_{1i}\hat{\alpha}_i. \end{cases} \tag{65}$$

Now averaging over r.v. $\acute{S}_i$, we have,

$$\begin{aligned} E\{Z_i\} &= E_{\acute{s}}\left(E_w\{Z_i\}\right) \\ &= \frac{1}{\acute{\beta}_i}\left(e^{-2\acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i} + 2\acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i - 1 \right), \end{aligned} \tag{66}$$

$$\text{Var}\{Z_i\} = E_{\acute{s}}\left(\text{Var}_w\{Z_i\}\right) + \text{Var}_{\acute{s}}\left(E_w\{Z_i\}\right) \tag{67}$$
$$= \frac{1}{\acute{\beta}_i^2}\left(3 - e^{-4\acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i} - 2e^{-2\acute{\beta}_i\hat{\eta}_{1i}\hat{\alpha}_i}\left(1 + 4\acute{\beta}_i\hat{\eta}_i\hat{\alpha}_i\right) \right).$$

Substituting $E\{Z_i\}$, and $\text{Var}\{Z_i\}$ in Eq. (61) and Eq. (61), we have,

$$E\{T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} = \sum_{i\in\zeta}\left(e^{-\frac{\hat{\eta}_{1i}2\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}} + \frac{2\hat{\eta}_{1i}\sqrt{2}}{\hat{\eta}_{2i}\lambda_i} - 1 \right), \tag{68}$$

$$\text{Var}\{T(\hat{\mathbf{V}}|\,\mathbf{s},\ \hat{\alpha},\ \hat{\eta}_1,\ \hat{\eta}_2)\} = \tag{69}$$
$$\sum_{i\in\zeta}\left(3 - e^{-\frac{\hat{\eta}_{1i}4\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}} - 2e^{-\frac{\hat{\eta}_{1i}2\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}}\left(1 + \frac{4\hat{\eta}_{1i}\sqrt{2}}{\hat{\eta}_{2i}\lambda_i}\right) \right).$$

Therefore, P_e for an ICA-based detector is given as,

$$P_{e\_ICA} = Q\left(\frac{|E\{T\}|}{\sqrt{\text{Var}\{T\}}} \right) \tag{70}$$

It can be observed from Eq. (70) that the decoding error probability of the ML decoder applied to the estimated watermark is a function of *WIR* and *SWR*. The performance of the proposed ICA-based detector given by Eq. (70) for different values of *WIR* and *SWR* is plotted in Fig. 5. It can observed from Fig. 5 that the proposed ICA-based detector perform superior than the detector operating without host-interference cancelation.

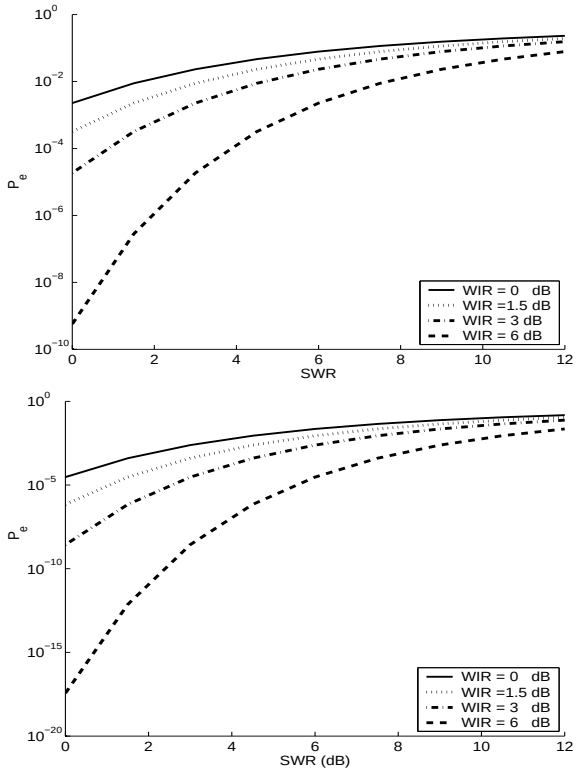

Figure 5: The P_e performance of the Proposed ICA-based Detector for different values of *WIR* and one bit per $|\zeta|$ Coefficients Embedding, i.e. $|\zeta| = 5$ (top) , $|\zeta| = 10$ (bottom) (theoretical values)

4.4 Maximum Watermarking-Rate: *Performance Analysis*

Maximum watermarking-rate (MWR) is another watermarking performance measure which indirectly depends on

the detector structure. Researchers in data hiding community have proposed various host interference suppression methods based on linear as well as non-liner filtering to improve MWR performance of a blind detector. For example, Su et al in [44] used optimal linear filtering to suppress host interference at the blind detector to improve MWR. The MWR performance of the proposed ICA-based watermark detector is evaluated for one bit per coefficient embedding case, i.e, $X_0 = S_0 + \alpha_0 W_0 b$. Let us assume the the received watermarked sample is corrupted by independent additive white Gaussian noise, with mean zero and variance $\sigma^2_{n_0}$. Here using CLT $\tilde{X}_0$ can be approximated by a Gaussian r.v. with mean zero and variance

$$\sigma^2_{\tilde{x}_0} = \sigma^2_{s_0} + \sigma^2_{v_0} + \sigma^2_{n_0}. \tag{71}$$

In this case, the estimated watermark sample, $\hat{V}_0$ can be expressed as,

$$\hat{V}_0 = \eta_{10}\alpha_0 W_0 + \eta_{20} S_0 + N_0 \tag{72}$$

Again $\hat{V}_0$ can also be approximated by Gaussian r.v. with mean zero and variance

$$\sigma^2_{\hat{v}_0} = \eta^2_{10}\sigma^2_{v_0} + \eta^2_{20}\sigma^2_{s_0} + \sigma^2_{n_0}. \tag{73}$$

The MWR of watermarking schemes based on additive embedding using blind correlation-based watermark detector can be approximated by the capacity of an additive white Gaussian noise channel, i.e.,

$$R_{Cor} = \frac{1}{2}\log_2\left(1 + \frac{\sigma^2_{v_0}}{\sigma^2_{s_0} + \sigma^2_{n_0}}\right) \tag{74}$$

Similarly, MWR using an informed detector can be expressed as,

$$R_{Informed} = \frac{1}{2}\log_2\left(1 + \frac{\sigma^2_{v_0}}{\sigma^2_{s_0}}\right) \tag{75}$$

And, MWR of the proposed ICA-based watermark detector is given as,

$$R_{ICA} = \frac{1}{2}\log_2\left(1 + \frac{\hat{\eta}^2_{10}\sigma^2_{v_0}}{\hat{\eta}^2_{20}\sigma^2_{s_0} + \sigma^2_{s_0}}\right) \tag{76}$$

Since the ICA scheme used for watermark estimation has reasonably good source separation performance [34, 35], therefore following inequality will hold,

$$\frac{\hat{\eta}^2_{10}\sigma^2_{v_0}}{\hat{\eta}^2_{20}\sigma^2_{s_0} + \sigma^2_{n_0}} \leq \frac{\sigma^2_{v_0}}{\sigma^2_{s_0} + \sigma^2_{n_0}} \tag{77}$$

$$\Rightarrow R_{ICA} \geq R_{Cor} \tag{78}$$

It can be observed from Eq. (78) that the proposed ICA-based detector performs better than the blind detector operating without suppressing the host signal interference. In addition, MWR performance of ICA-based detector is bounded below by the blind detector (0% suppression) and bounded above by an informed detector (100% suppression).

Performance analysis of the proposed ICA-based watermark detector indicates that it performs better than existing blind watermark detectors [1, 2, 3, 4, 5] operating without reducing host signal interference. This improved detection performance of ICA-based detector can be attributed to its host signal interference suppression at the detector.

4.5 Estimation of Masking Threshold, Distribution Parameter and WIR factor

This section provides details on how to estimate masking threshold, $\hat{\alpha}$, host distribution parameter, $\hat{\beta}$, and $\hat{\eta}_1$, $\hat{\eta}_2$ at the blind detector. The $\tilde{\mathbf{x}}$ is analyzed at the blind detector to estimate $\hat{\alpha}$ based on HAS. It is reasonable to assume that $\hat{\alpha}$ estimated from watermarked audio is similar to the $\hat{\alpha}$ from the corresponding an unwatermarked audio clip given that embedding and attack-channel distortion are imperceptible. To validate this assumption, we estimated $\hat{\alpha}$ from both the unwatermarked and corresponding watermarked music clips. To this end four music clips (*Pos1, Pop2, Classic*, and *Vocal*) listed in Table 1) were used. Here music clips *Pop1* and *Classic* were watermarked using FSSS based watermarking scheme proposed in [5] and *Pop2* and *Vocal*, were watermarking using audio watermarking scheme presented in [3]. Plots of the $\hat{\alpha}$ estimated from the each watermarked music clip, $\hat{\alpha}_W$ and corresponding unwatermarked music clip $\hat{\alpha}_{UW}$ are given in Fig. 6.

It can be observed from Fig. 6 that for both embedding schemes $\hat{\alpha}_W \approx \hat{\alpha}_{UW}$. Similarity between $\hat{\alpha}_W$ and $\hat{\alpha}_{UW}$, for the music clips listed in Table 1, in terms of mean squared error (MSE) (in dB) is {*Pos1, Melodic, Pop2, Classic, Vocal*} = {0.21566, 1.7321, 2.4507 , 1.7716, 0.21566}. Here watermarked music clips were generated using FSSS-based watermarking. These results shows that it is reasonable to estimated masking threshold from the watermarked audio at the blind detector.

Distribution parameter, β, can be estimated from the estimated variance $\hat{\sigma}^2_s$ of the host audio, which can be estimated from the watermarked audio available at the detector

$$\hat{\sigma}^2_s = \hat{\sigma}^2_x - \frac{1}{M}\sum_j \hat{\alpha}^2_j \tag{79}$$

where $\hat{\alpha}^2_m$ is the variance of the watermark sequence for m^{th} audio segment and M is total number of watermarked segments.

Here $\hat{\sigma}^2_x$ is estimated using sample variance, i.e.,

$$\hat{\sigma}^2_x = \frac{1}{M}\sum_j \mathbf{X}^2{}_j - \frac{1}{M^2}\left(\sum_j \mathbf{X}_j\right)^2 \tag{80}$$

It is important to mentions that if this estimate is used to calculate sufficient statistics, this will introduce additional dependence between watermark and sufficient statistics which is hard to analyze theoretically. Due to this

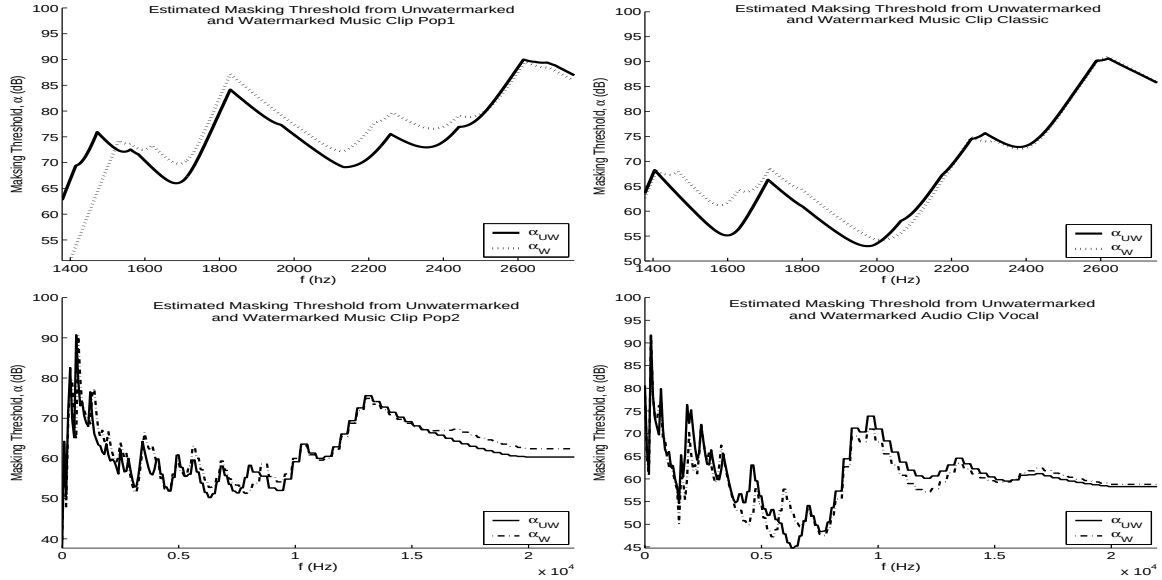

Figure 6: Plots of the estimated masking threshold from watermarked music clips $\hat{\alpha}_W$ and unwatermarked music clips $\hat{\alpha}_{UW}$

added dependence, slight variation between theoretical approximation and experimental results is expected.

The problem of estimating η_1 and η_2 is bit hard due to ambiguity in the scale and sign of the estimated sources using ICA. However, if we assume that scale and sign ambiguity of the separated sources is resolved and *WIR* factor $D^{\text{interf}} = \frac{\eta_1}{\eta_2}$ is known then, using Eq. (23) and (24), η_1 and η_2 can be estimated by simultaneously solving the following expressions,

$$\hat{\sigma}_{\hat{v}}^2 = \eta_1^2\hat{\alpha}^2 + \eta_2^2\hat{\sigma}_s^2, \eta_1^2 = D^{\text{interf}}\eta_2^2. \tag{81}$$

Here D^{interf} can calculated using separation quality measure of the ICA method used as discussed in [34, 35].

5 Simulation Results

This section provides detection performance of the proposed ICA-based watermark detector (ICAWD) and its comparison with the conventional normalized correlation watermark detector (NCWD) [1]. The proposed ICAWD can be used to detect watermark for almost all existing SS-based watermark embedding schemes [1, 2, 3, 5, 4]. However detection performance of the proposed detector is compared with Swanson et al's SS-based audio watermarking scheme [3]. Swanson et al's [3] proposed scheme used correlation based detector for watermark detection. To provide a fair performance comparison of both the proposed ICAWD and the NCWD, the proposed ICAWD is used in the estimation-correlation-based detection settings. The simulation results presented based on FSSS-based audio watermark embedding scheme presented in [5]. Detains of watermark embedding using FSSS [5] outlined here.

5.1 FSSS-based Watermark Embedding

The block diagram of the FSSS-based watermark generation and embedding used for simulations is illustrated in Fig. 7. The watermark is generated using a pseudo-random noise generator obeying non-Gaussian distribution to satisfy the non-Gaussianity requirement of the ICA model. A secret key $\mathrm{K_w}$ is used as a '*seed*' for the pseudo-random noise generator for watermark generation. In addition, same watermark is embedded in two consecutive audio segments, i.e., if watermark $\mathbf{V}$ is embedded into i^{th} audio segment then same watermark is also embedded into $(i+1)^{th}$ segment. Repeated embedding is a necessary condition of the proposed detector to separate hidden signals obeying heavy-tail distribution, especially for BSS from underdetermined linear mixtures [40, 16]. For audio watermarking using FSSS, a secret key, K_{sb} is used to select subband from watermark embedding.

5.2 Watermark Detection

The proposed modified ICA-based detector has access to the secret key K only, which is combination of K_{sb} and $\mathrm{K_w}$, i.e., $\mathrm{K} = \mathrm{K}_{sb}|\mathrm{K_w}$. The watermark detection process for FSSS-based audio watermarking under proposed detection scheme consists of watermark estimation using ICA framework followed by correlation based detection. The main steps of the detection process are outlined below:

- ***Sync Point Extraction***: The received audio signal is analyzed first to extract the set of sync points (SP) [4, 5] used to combat desynchronization attacks.
- ***Segmentation***: An audio frame consisting of n-samples is selected around each $SP_i : i = 1, 2 \cdots M$. Where M is cardinality of SP set.

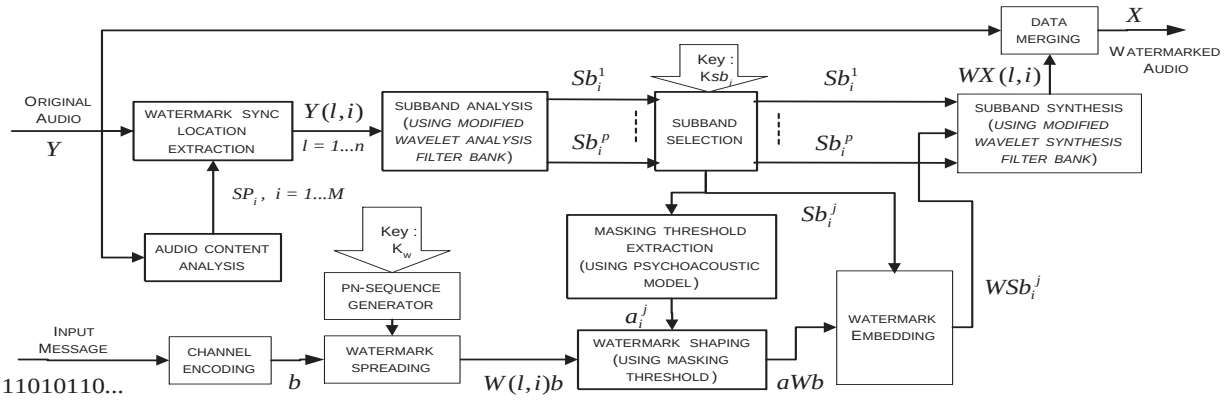

Figure 7: Block diagram of the FSSS-based watermark embedding

– ***Frame Decomposition***: Each frame is then decomposed into p–subband signals using l–level analysis filter bank described in [5] .

– ***Subband Selection***: A secret key K_{sb_i}, is used to select a subband from lower $(p-1)$–subbands of i^{th} and $(i+1)^{th}$ frame i.e. $\widetilde{\mathbf{Sb}}_i^j$, and $\widetilde{\mathbf{Sb}}_{i+1}^j$.

– ***Watermark Estimation***: The selected subband signals, i.e. $\widetilde{\mathbf{Sb}}_i^j$ and $\widetilde{\mathbf{Sb}}_{i+1}^j$ are used to estimated the embedded watermark, $\mathbf{V}$ Here, observation matrix, $\mathbf{X}$, can be expressed as, $\mathbf{X} = \left[\widetilde{\mathbf{Sb}}_i^j, \widetilde{\mathbf{Sb}}_{i+1}^j\right]^T$.

Existing BSS schemes for underdetermined mixtures based on ICA model [27, 39] to estimate watermark from the watermarked image for the proposed detector. However, in this paper, the proposed ICAWD uses the statistical ICA using mean-field approaches presented in [39] for watermark estimation from the watermarked audio. The watermark detection stage uses the correlation based similarity measure to determine the presence or the absence of the embedded watermark from the estimated sources. It is important to mention that permutation ambiguity in the estimated sources using ICA will contribute nonzero P_e due to incorrect source decoding. The error due to ambiguity in the permutation of the estimated sources is reduced by adding correlation based watermark detection (resp. decoding). However for the sake of simplicity, during analysis part in Section 4, error due to incorrect source decoding is neglected here.

– ***Information Decoding***: A binary hypothesis test is used to determine the presence or the absence of the embedded watermark in the estimated signal. For fast and reliable information decoding, normalized correlation between the estimated watermark and the key dependent watermark generated at the watermark detector are used. The normalized correlation is then compared against decision threshold, *Th*, to determine the presence or the absence of watermark. Following binary hypothesis test is used to decode binary information,

$$H_1: \quad \max \left| ncor\left(\hat{\mathbf{S}}^{(r)}, \mathbf{W}^{(q)}\right)\right| \geq Th \text{ Decode } q$$
$$H_0: \quad \text{otherwise } \textit{no watermark}$$

where *ncor*(.,.) is the normalized correlation function defined as:

$$ncor\left(\hat{\mathbf{S}}^{(r)}, \mathbf{W}^{(q)}\right) = \frac{\sum_{l=-n}^{n} \hat{S}_l^{(r)} W_{t+l}^{(q)}}{\sqrt{\sum_{l=0}^{n}\left(\hat{S}_l^{(r)}\right)^2 \sum_{l=0}^{n}\left(W_l^{(q)}\right)^2}} \tag{82}$$

where $\hat{\mathbf{S}}^{(r)}$ is the estimated signals using ICA, *Th* is the decision threshold (for our simulation results Th was set to 0.15, which corresponds to false positive rate, $P_{fp} = 10^{-4}$), $r = 1, 2, 3$, and $q \in \{0, 1\}$.

5.3 Experimental Results

To evaluate the robustness performance of the proposed ICAWD, several experimental tests were performed in which the watermarked audio is subjected to commonly encountered degradations. These degradations include addition of white and colored noise, resampling, lossy compression (MP3 Audio compression), filtering, time- and frequency-scaling, and stirmark benchmark attacks for audio [18, 20].

Decoding error probability, Pb_e, at the watermark detector is used for performance evaluation. Here Pb_e is defined as,

$$Pb_e = \left(1 - \frac{N_d}{N_e}\right) \tag{83}$$

where N_d is number of bits correctly detected and N_e number of bits embedded into the audio clip.

Block diagram of the proposed ICAWD and the traditional correlation based detector e.g. NCWD used for FSSS audio watermark detection process is given in Fig. 8. The watermark detector given in Fig. 8 acts as ICAWD when switch S is connected to terminal 1 and NCWD when S is connected to 2.

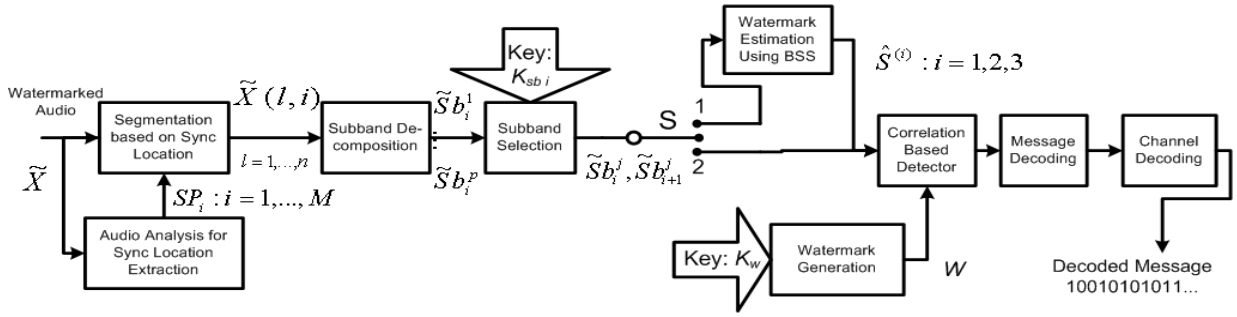

Figure 8: Block diagram of the ICAWD and the NCWD used for performance comparison

5.4 Robustness Performance

To evaluate the robustness performance of the proposed watermarking scheme we have performed several experimental tests in which the watermarked audio is subjected to commonly encountered degradations. These degradations include addition of white and colored noise, resampling, lossy compression (MPEG audio compression), filtering, time- and frequency-scaling, multiple watermarking, and StirMark benchmark attacks for audio.

The robustness performance of the proposed scheme against common degradations for the above settings is discussed next.

5.5 Data Set

Experimental results presented here are based on the data set consisting of the *sound quality assessment material* (SQAM) audio database downloaded from [45] and five audio clips listed in Table 1. All audio clips used for the performance evaluation here are based on mono audio channel sampled at 44.1 kHz with 16 bits resolution.

In our experiments, the watermarks are generated and embedded using FSSS-based audio watermarking scheme presented in [5]. A perceptual mask is estimated using method discussed in [5]. This mask is then multiplied by 200 independently generated pseudo-random sequences $\mathbf{W}$, with zero-mean and unit variance, to generate 200 independent watermarks. In case of ICAWD, the pseudo-random sequences, $\mathbf{W}$, follow Laplacian distribution, i.e.,

$$f_W(\tau) = \frac{\beta}{2} e^{-\beta|\tau|}, \quad |\tau| < \infty \tag{84}$$

where $\beta = \frac{\sqrt{2}}{\sigma_{\mathbf{W}}}$, and for the NCWD $\mathbf{W}$ follows normal distribution. These 200 random watermarks are embedded in each audio clip according to Eq. (1) that resulted 4000 watermarked audio clips. Experimental results presented in the following sections are averaged over 4000 watermarked audio clips.

5.6 Parameter Settings

Simulation results presented in this section are based on the following system settings:

- Salient point list (SP) was assumed to be available at the detector, therefore decoding bit error probability P_e presented here is due decoding bit error only.
- Audio frame size ($2^l N_1$) was set to 2^{13} for $f_s = 44.1$ kHz.
- Five-level wavelet decomposition was used, i.e. $l = 5$, therefore eight target subbands were available for watermark embedding.
- Only one subband was selected at random from eight target subbands for watermark embedding (except multiple watermark embedding case).
- Target false positive rate P_{fp} was set to 3.5×10^{-4} which corresponds to decoding threshold $Th = 0.15$ (using Eq. (42)).
- False positive bit rate, P_{fp}, was calculated by applying original (unwatermarked) music clip the proposed detector, and average false positive for the the 20 audio clips used for performance evaluation was calculated to be 2.9×10^{-4}.
- Robustness performance in terms of average decoding bit error rate was calculated without channel coding.
- In case of ICAWD, watermark repeating factor of two was used during watermark embedding process, i.e., two consecutive audio frames were watermarked with same watermark $\mathbf{w}$.

The above settings for watermark embedding using FSSS-based audio watermarking yielded *per sample embedding capacity* of 1 bit per 512 sample.

Fidelity (or transparency) performance of the embedded watermark is evaluated based on the objective degradation measure. Signal-to-watermark ratio (SWR) is used for the objective degradation here which is calculated as,

Table 1: Audio Clips used for Performance Evaluation

Singer Name, Song Title	***Genre***	***Duration*** (sec)
Back Street Boys, *I Want It That Way*$\cdots$	Pop, (Pop1)	22
L. Mangeshkar, *Kuch Na Kaho*$\cdots$	Melodic, (Melodic) (Melodic)	15
A. Bhosle, & R. Sharma, *Kahin Aag Laga*$\cdots$	Pop, (Pop2) (Pop2)	10
N. F. A. Khan, *Afreen Afreen*$\cdots$	Semi-Classic, (Classical)	20
Suzanne Vega, *Tom's diner*$\cdots$	Female Vocal, (Spoken Language)	5

$$SWR = 10\log_{10}\left(\frac{\sigma_s^2}{\sigma_\nu^2}\right) \quad (85)$$

where σ_ν^2 is calculated using Eq. (3).

The average SWR the watermark audio clips used for simulation was $Ave_{SWR} = 42.7$ (dB), $\sigma_{SWR} = 9.17$, $max_{SWR} = 74.5$ (dB), and $min_{SWR} = 21.5$ (dB). Calculated SWR from watermarked audio clips indicates that on the embedded watermark is very weak compared to the original audio.

5.7 Detection Performance

Detection performance of the proposed detector is evaluated for various audio degradations. Detection of the proposed ICAWD and its comparison with NCWD for each degradation is provided next.

5.7.1 Addition of White Noise

: White Gaussian noise ranging from zero to 200 % of the power of the audio signal was added to the corresponding watermarked audio clips. The P_e average over 4000 watermarked audio clips for ICAWD and NCWD for different SNR values are plotted in Fig. 9 which shows that the ICAWD performs better than the NCWD. Superior detection performance of ICAWD than the NCWD can be attributed to its host signal interference cancellation capability. It can be observed from Fig. 9 that for SS-based watermarking very low decoding bit error probability is achievable even in the presence of noise with 60 - 70 % power of the audio signal.

5.7.2 Resampling

To simulate resampling attack, a watermarked audio signal was down-sampled at a sampling rate of $\frac{f_s}{r_f}$ (where r_f denotes resampling factor) and then interpolated back to f_s. The watermark detection was then applied to the resulting watermarked audio clips. Average P_e for $r_f = 2, \cdots 10$, is given in Fig. 10 which shows that the proposed watermarking scheme (using ICAWD) can withstand resampling attacks with r_f value up to 5 for each watermarked audio clip, similar decoding performance is achievable for NCWD by using channel coding. Again ICAWD performs better than the NCWD and its superior detection performance can be attributed to its host signal interference suppression capability.

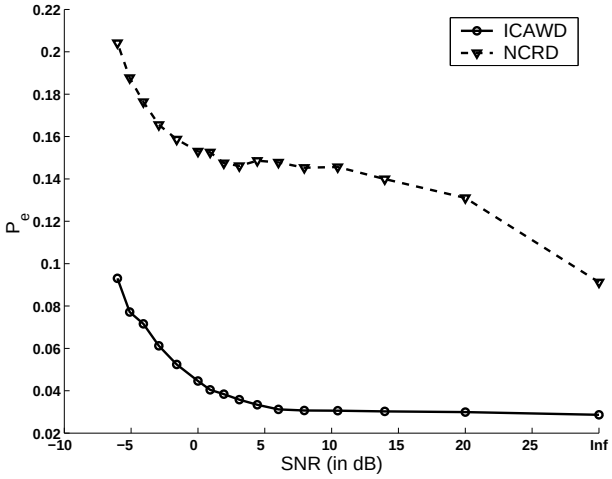

Figure 9: Detection performance comparison for AWGN attack.

5.7.3 Lossy Compression

Lossy compression for audio (e.g. MP3) is generally applied to the digital audio for multimedia applications like transmission and storage to reduce the bit rate. To test the survivability of the watermark, audio encoding/decoding was applied to the watermarked audio using ISO/MPEG-1 Audio Layer III [47] coder at bit rates 32, 64, 96, 112, 128, 192, 256, and 320 k bits/s (kbps). The average P_e for lossy compression attacks for bit rates rates 32, 64, 96, 112, 128, 192, 256, and 320 (kbps) is given in Fig. 11. It has been observed from Fig. 11 that the detection per-

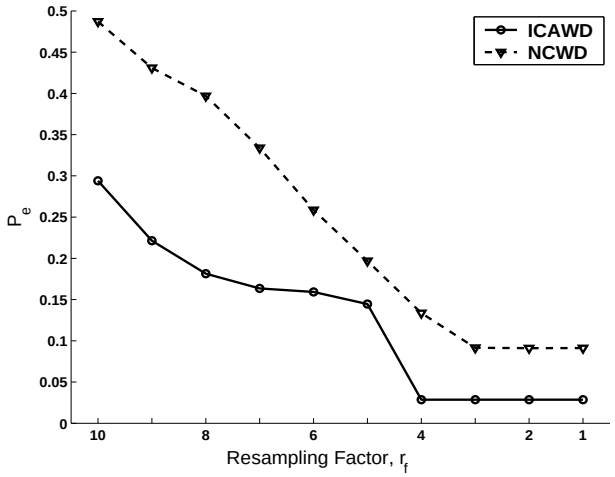

Figure 10: Detection performance comparison for resampling attack.

formance for both detectors deteriorates as the bit rate of the encoder/decoder decreases; this is due to the stronger distortion introduced by the encoder for lower bit rates. In addition, the ICAWD performs better than the NCWD.

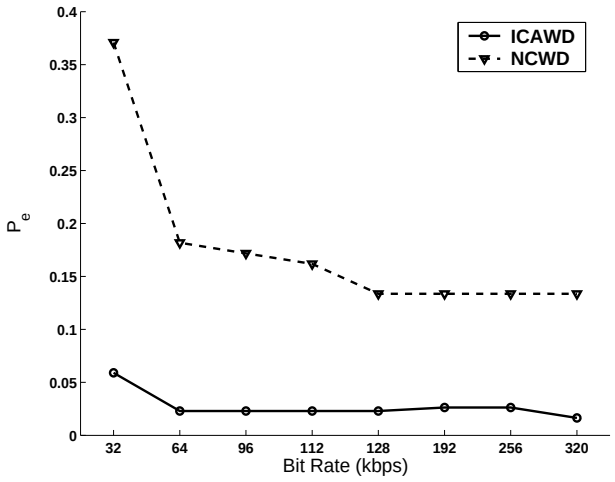

Figure 11: Detection performance comparison for MP3 compression attack

5.7.4 Addition of Colored Noise

To simulate an attack with colored noise, white Gaussian noise was spectrally shaped according to the estimated masking threshold using corresponding watermarked audio clip based on the HAS model [46, 47]. This just audible colored noise was then added to the watermarked audio signal. Average P_e for the resulting watermarked audio clips is presented in Fig. 12. It has been observed from Fig. 12 that NCWD performs poorly, this is due to increase in interference level, as the colored noised is generated with a process almost identically to that of the watermark generation. Therefore, additive colored noise acts as a second watermark interfering with the watermark to be detected. On the other hand, ICAWD is efficient in handling such attacks due to its interference cancellation ability.

5.7.5 Rescaling

Rescaling attacks include time- and frequency-scaling. Time-scaling attacks can be used to desynchronize a watermark detector for SS-based watermarking systems. To test the robustness of the proposed scheme against time-scaling attacks, the watermarked audio clips were time-scaled with time-scaling factor, TSp(n) = +(-) 1 %. The detection performance for time-scaling attack using both detection schemes, e.g., ICAWD and NCWD is given in Fig. 12.

The frequency-scaling attacks are generally used to deteriorate the detection performance of the frequency domain watermarking schemes. As the proposed watermarking scheme is also a frequency domain watermarking scheme; therefore, it is reasonable to test the robustness performance of the proposed scheme against frequency-scaling attacks as well. To simulate frequency-scaling attack, the watermarked audio clips were frequency-scaled using frequency-scaling factor, FSp(n) = +(-) 1%. The detection performance for the resulting audio clips for both detection schemes, e.g., ICAWD and NCWD is presented in Fig. 12. It can be observed from Fig. 12 that the proposed scheme can withstand rescaling attack of TS $\leq \pm 1\%$ and FS$\leq \pm 1\%$ (especially for ICAWD).

5.7.6 Filtering

To test the robustness of the proposed watermarking scheme against filtering attacks, the watermarked audio signals were subjected to lowpass filtering (LPF), highpass filtering (HPF), and bandpass filtering (BPF) attacks. The specification of filters used for the filtering attacks are,

1. Lowpass Filter: cut-off frequency: $f_c = 5$ kHz with 12 dB/octave roll-off
2. Highpass Filter: cut-off frequency: $f_c = 1000$ Hz with 12 dB/octave roll-off
3. Bandpass Filter: cut-off frequencies: $f_{clow} = 50$ Hz, and $f_{cup} = 5.5$ kHz with 12 dB/octave roll-off

Detection performance comparison of the ICAWD and the NCWD for LPF, HPF, and BPF attacks is given in Fig. 12.

5.7.7 StirMark Audio Benchmark Attacks

For StirMark for audio benchmark attack, watermarked audio clips were subjected to StirMark audio benchmark attacks. The StirMark audio benchmark software, available at [20], was used in the default parameters settings. The decoding bit error probability, P_e, averaged over 100 watermarked audio clips with the ICAWD and the NCWD, is given in Table 2. It can be observed from Table 2 that

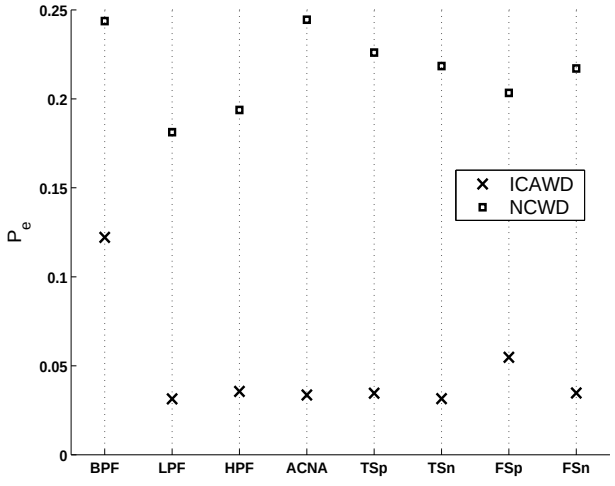

Figure 12: Detection performance comparison between the ICAWD and the NCWD for filtering (LPF, HPF, BPF), rescaling (TSp, TSn, FSp, FSn), requantization (Res), and colored noise addition (ACNA) attacks

the proposed ICAWD based scheme using exhibits superior detection performance than the NCWD. Better performance of ICAWD can be attributed to its better host signal suppression capability.

6 Conclusion

An improved watermark detector for additive embedding is presented here. The proposed watermark detector is capable of canceling the host-signal interference at the watermark detector. Bind watermark detection, lower host-signal interference at the detector, improved decoding, detections and watermarking-rate performances are the salient features of the proposed ICAWD. The proposed ICAWD can be used for SS-based watermarking for all types of multimedia data, e.g., audio, video, images, etc. The theoretical results show that the proposed detector performs significantly better than existing blind detectors. Simulation results for real-world data show that the proposed ICAWD performs much better than the traditional NCWD. Moreover, the detection performance of the proposed detector can be improved further by employing channel coding. It is important to mention that better detection performance of ICAWD comes at the cost of security, as ICAWD requires repeated embedding (at least twice) which makes embedded watermark more vulnerable to watermark estimation attacks than without repeated embedding.

References

[1] Cox, I.J., Miller, M.L., and Bloom, J.A.:(2001) *Digital Watermarking*, Morgan Kaufmann, San Francisco.

Table 2: Performance Comparison for StirMark Audio Benchmark Attacks

Decoding Bit Error Probability, P_e		
StirMark Attack NCWD	ICAWD	
addbrumm_100	0.088	0.0091
addbrumm_1100	0.088	0.0091
addbrumm_2100	0.088	0.0091
addbrumm_3100	0.1023	0.0091
addbrumm_4100	0.1257	0.0091
addbrumm_5100	0.1412	0.0091
addbrumm_6100	0.1477	0.0091
addbrumm_7100	0.1904	0.0091
addbrumm_8100	0.2228	0.0091
addbrumm_9100	0.2293	0.0234
addbrumm_10100	0.2293	0.0491
addfftnoise	1	1
addnoise_100	0.088	0.0491
addnoise_300	0.088	0.0491
addnoise_500	0.088	0.0491
addnoise_700	0.088	0.0634
addnoise_900	0.088	0.0634
addsinus	0.088	0.0634
amplify	0.088	0.0491
compressor	0.088	0.0491
copysamples	0.529	0.1749
cutsamples	0.791	0.4835
dynnoise	0.1056	0.0667
echo	0.0818	0.0667
exchange	0.1056	0.0818
extrastereo_30	0.1056	0.0818
extrastereo_50	0.1056	0.0818
extrastereo_70	0.1056	0.0818
fft_hlpass	0.1074	0
fft_invert	0.1056	0.0818
fft_real_reverse	0.1056	0.0818
fft_stat1	0.1295	0.0238
fft_test	0.1056	0.0238
flippsample	0.1281	0.0725
invert	0.088	0.0491
lsbzero	0.1056	0.0818
normalize	0.088	0.0673
rc_highpass	0.0945	0.0491
rc_lowpass	0.088	0
smooth	1	1
resample	0.1056	0
smooth2	0.1056	0
stat1	0.1056	0
stat2	0.1056	0.0818
voiceremove	1	1
zerocross	0.088	0
zeroremove	0.2759	0.0363
zerolength	0.2189	0.0607

[2] Cox, I.J., Kilian, J., Leighton, T., and Shamoon, T.:(1997) Secure Spread Spectrum Watermarking for Multimedia, *IEEE Trans. on Image Processing*, vol. 6(12), pp. 1673–1687.

[3] Swanson, M.D., Zhu, B., Tewfik, A.H., and Boney, L.:(1998) Robust Audio Watermarking using Perceptual Masking, *Signal Processing*, vol. 66(3), pp. 337–355.

[4] Wu, C.-P., Su, P.-C., and Kuo, C.-C. J.:(1999) Robust Audio Watermarking for Copyright Protection, *Proc. SPIE's* 44^{th} *Ann. Meet. Adv. Sig. Proc. Alg. Arch. Impl. IX (SD39)*, vol. 3807, pp. 387–397.

[5] Malik, H., Khokhar, A., and Ansari, R.:(2004) Robust Audio Watermarking using Frequency Selective Spread Spectrum Theory, *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP'04)*, Montreal, Canada, pp. 385–388.

[6] P-Gonzalez, F., Balado, F., and Hernández, J.R.:(2003) Performance Analysis of Existing and new Methods for Data Hiding with Known-Host Information in Additive Channels, *IEEE Trans. on Signal Processing*, vol. 51(4), pp. 960–980.

[7] Hernandez, J., Amado, M., and Perez-Gonzalez, F.:(2000) DCT-domain watermarking techniques for still images: Detector performance analysis and a new structure, *IEEE Trans. on Image Processing*, vol. 9(1), pp. 55–68.

[8] Briassouli, A., and Strintzis, M.:(2004) Locally Optimum Nonlinearities for DCT Watermark Detection, *IEEE Trans. on Image Processing*, vol. 13(12), pp. 1604-1617.

[9] Noel, S., and Szu, H.:(2000) Multimedia Authenticity with ICA Watermarks, *IS&T/SPIE. Proc., Wavelet Applications VII*, vol. 4056, pp. 175–184.

[10] Serrano, F., and Fuentes, J.:(2001) Independent Component Analysis Applied to Digital Image Watermarking, *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'01)*, vol. 3, pp. 1997-2000.

[11] Toch, B., Lowe, D., and D. Saad, D.:(2003) Watermarking of Audio Signals using using Independent Component Analysis, *Proc.* 3^{rd} *Int. Conf. WEB Delivering of Music*, pp. 71-74.

[12] Sener, S., and Gunsel, B.:(2004) Blind Audio Watermark Decoding using Independent Component Analysis, *Proc.* 17^{th} *Int. Conf. Patt. Reco. (ICPR'04)*, vol. 2, pp. 875–878.

[13] Bounkong, S., Toch, B., Saad, D., and Lowe, D.:(2002) ICA for Watermarking Digital Images, *J. Machine Learning Research 1*, pp. 1–25.

[14] Yu, D., Sattar, F., and Ma, K.:(2002), Watermark Detection and Extraction using Independent Component Analysis, *EURASIP J. Applied Signal Processing*, pp. 92–104.

[15] Malik, H., Khokhar, A., and Ansari, R.:(2005) Improved Watermark Detection for Spread-Spectrum based Watermarking using Independent Component Analysis, *Proc.* 5^{th} *ACM Workshop On Digital Rights Management (DRM'05)*, Washington DC, pp. 102–111.

[16] Malik, H., Khokhar, A., and Ansari, R.:(2006) New detector for spread-spectrum based image watermarking using underdetermined ICA, *IS&T/SPIE Conf. Security, Steganography, and Watermarking of Multimedia Contents VIII '06*, vol. 6072, pp. 747–758.

[17] Malik, H., Khokhar, A., and Ansari, R.:(2006) Blind Detection for Additive Embedding Using Underdetermined ICA, *Proc.* 8^{th} *IEEE Int. Symposium on Multimedia, (ISMapos'06)*, pp. 758–761.

[18] Steinebach, M., Lang, A., Dittmann, J., and Prtitcolas, F.A.P.:(2002) StirMark Benchmark: Audio Watermarking Attacks based on Lossy Compression, *Proc. SPIE Security Watermarking Multimedia*, vol. 4675, pp. 79–90.

[19] Lang, A., Dittmann, J., Spring, R., and Vielhauer, C.:(2005), Audio watermark attacks : from single to profile attacks, *Proc. ACM Multimedia and Security Workshop (MM & Sec'05)*, New York, NY, USA, pp. 39–50.

[20] *StirMark Benchmark for Audio*, available at `http://amsl-smb.cs.uni-magdeburg.de/smfa/main.php`, accessed on June 23, 2008.

[21] Bell, A., and Sejnowski, T.:(1995) An Information Maximisation Approach to Blind Separation and Blind Deconvolution, *Neural Computation*, MIT Press Journals, vol. 7(6), pp. 1129–1159.

[22] Hyvarinen, A., Karhunen, J., and Oja E.:(2001) *Independent Component Analysis*, John Wiley & Sons.

[23] Comon, p.:(1994) Independent Component Analysis, A New Concept?, *EURASIP Signal Processing*, vol. 36(3), pp. 287–314.

[24] Cao, X.-R., and Liu, R.-W.:(1996) Genaral Approach to Blind Source Separation, *IEEE Trans. Signal Processing*, vol. 44(3), pp. 562–571.

[25] Eriksson, J., and Koivunen, V.:(2004) Identifiability, Separability, and Uniqueness of Linear ICA Models, *IEEE Signal Processing Letters*, vol. 11(7), pp.601–604.

[26] Davis, M.:(2004) Identifiability Issues in Noisy ICA, *IEEE Signal Processing Letters*, vol. 11(5), pp. 601–604.

[27] De Lathauwer, L., Comon, P., De Moor, B., and Vandewalle, J.:(1999) ICA Algorithms for 3 Sources and 2 Sensors, *Proc. IEEE Signal Processing Workshop Higher-Order Statistics (HOS'99)*, pp. 116–120.

[28] Gaeta, M., and Lacoume, J.-L.:(1990) Source Separation Without Prior Knowledge: The Maximum Likelihood Solution, *Proc. EUSIPCO'90*, pp. 621–624.

[29] Moulinesand, E., Cardoso, J-F., and Gassiat, E.:(1997) Maximum Likelihood for Blind Separation and Deconvolution of Noisy Signals using Mixture Models, *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'97)*, pp. 3617–3620.

[30] Cardoso, J.-F.:(1999) High-order Contrasts for Independent Component Analysis, *Nural Computation*, Elsevier Science, vol. 11(1), pp. 157–192.

[31] Bofill, P., and Zibulevsky, M.:(2001) Underdetermined Blind Source Separation using Sparse Representations, *EURASIP Signal Processing*, vol. 81(11), pp. 2353–2362.

[32] Zibulevsky, M., and Zeevi, Y.Y.:(2002), Extraction of a Single Source from Multichannel Data using Sparse Decomposition, *Neurocomputing*, Elsevier Science, vol. 49, pp. 163–173.

[33] Li, Y., Cichocki, A., and Amari, S.:(2004), Analysis of Sparse Representation and Blind Source Separation, *Neural Computation*, MIT Press Journals, vol. 16(6), pp. 1193–1234.

[34] Gribonval, R., Benaroya, L., Vincent, E., and Fevotte, C.:(2003) Proposals for Performance Measurement in Source Separation, *Proc.* 4^{th} *Int. Sym. Independent Component Analysis and Blind Source Separation*, pp. 763–768.

[35] Li, Y., Powers, D., and Peach, J.:(2000) Comparison of Blind Source Separation Algorithms, *Advances in Neural Networks and Applicaitons*, N. Mastorakis (Ed.), WSES, pp. 18–21.

[36] Cichocki, A., Douglas, S., and Amari, S.:(1998) Robust Techniques for Independent Component Analysis (ICA) with Noisy Data, *Neurocomputing*, Elsevier Science, vol. 22, pp. 113–129.

[37] Pajunen, P.:(1997) Blind Separation of Binary Sources With Less Sensors Than Sources, *Proc. Int. Conf. on Neural Networks (ICNN'97)*, vol. 3, pp. 1994–1997.

[38] Hyvarinen, A.:(1999) Fast Independent Component Analysis with Noisy Data using Gaussian Moments, *Proc. ISCS'99*.

[39] Hojen-Sorensen, P., Winther, O., and Hansen, L.K.:(2002) Mean-Field Approaches to Independent Component Analysis, *Neural Computation*, MIT Press Journals, vol. 14, pp. 889–918.

[40] Hansen, L.K., and Petersen, K.B.:(2003) Monoaural ICA of While Noise Mixture is Hard, *Proc. of Sym ICA and BSS (ICA2003)*, pp. 815–820.

[41] Poor, H. V.:(1994) *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 2^{nd}-ed.

[42] Kay, S.:(1998) *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice Hall, Upper Saddle River, New Jersey.

[43] Swanson, M., Zhu, B., and Tewfik, A.:(1996) Robust Data Hiding for Images, *Proc. IEEE Digital Signal Processing Workshop*, pp. 37–40.

[44] J. Su, J., Eggers, J., and Girod. B.:(2001) Analysis of Digital Watermarks Subjected to Optimum Linear Filtering and Additive Noise, *EURASIP Signal Processing*, vol. 81, pp. 1141–1175.

[45] *SQAM - Sound Quality Assessment Material*, `http://sound.media.mit.edu/mpeg4/audio/sqam/`, accessed on June 23, 2008.

[46] Zwicker, R. E., and Fastl, H.:(1999) *Psychoacoustics: Facts and Models*, Springer-Verlag, Berlin.

[47] Noll, P.:(1997) MPEG Digital Audio Coding, *IEEE Signal Processing Magazine*, vol. 14(5), pp. 59–81.

[48] Papoulis, A., and Pillai, S.:(2002) *Probability, Random Variables and Stochastic Processes*, McGraw-Hill, New York, 4^{th} Ed.

[49] Mallat, S.:(1989) A Theory for Multiresolution Signal Decomposition, the Wavelet Representation *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11(7), pp. 674 –693.

7 Appendix A: Statistical characterization of the wavelet coefficients of audio signals

To determine the statistical characterization of the subband coefficients of the real-world speech samples, the speech samples, $Y_i,\ i = 0, 1 \cdots n-1$ are assumed to be i.i.d. Laplacian random variable with mean zero and variance σ_y^2. The one-dimensional discrete wavelet transform (DWT) of audio signal, $\mathbf{Y}$, can be calculated using Mallat's algorithm [49]. The DWT coefficients using Mallat's algorithm [49], e.g., approximate coefficients a_k and detailed

coefficients d_k, at different scales can be expressed as,

$$a_i^{j-1} = \sum_{k=0}^{N_1-1} h_{k-2i} a_k^j \tag{86}$$

$$d_i^{j-1} = \sum_{k=0}^{N_1-1} g_{k-2i} a_k^j \tag{87}$$

where j denotes the resolution and i is the index.

Eq. (86) and (87) describe linear filtering operation using filters $\mathbf{h}$ and $\mathbf{g}$ followed by down-sampling. Here $\mathbf{h}$ and $\mathbf{g}$ are finite impulse response (FIR) quadrature-mirror filters, also known as the scaling and the wavelet filters, respectively. The scaling filter is a lowpass filter, while the wavelet filter is a highpass filter. Moreover, the top-level coefficients a^J represent the original signal $\mathbf{y}$. Eq. (86) and (87) can be expressed using a single equation,

$$S_i^{j-1} = \sum_{k=0}^{N_1-1} \delta_{k-2i} S_k^j \tag{88}$$

where δ_i is the weighting factor depending on the filter coefficients h_i and g_i, i.e. approximate coefficients or detailed coefficients, and S_i^j is the wavelet coefficient at j^{th}-level.

Here Eq. (88) states that a wavelet coefficient at an arbitrary level $j-1$, is a weighted sum of N_1 wavelet coefficients from j^{th}-level wavelet. The wavelet coefficients at $J-1$ level can be expressed as

$$S_i^{J-1} = \sum_{k=0}^{N_1-1} \delta_{k-2i} S_k^J \tag{89}$$

According to Eq. (88), each wavelet coefficient an arbitrary level $j : 1 \leq j \leq J-1$ is a weighted sum of i.i.d. r.v. (e.g. audio samples in our case), therefore, the pdf of a wavelet coefficient S_i^j at j^{th}-level, can be determined using joint characteristic function $\Phi_{s_i^j}(\omega)$. If we assume that audio sample Y_i, is a Laplacian r.v., then pdf of Y_i can be expressed as,

$$f_y(\tau) = \frac{\gamma}{2} e^{-\gamma|\tau|}, \ |\tau| < \infty \tag{90}$$

where $\gamma = \frac{\sqrt{2}}{\sigma_y^2}$

Here characteristic function of r.v. Y_i, $\Phi_{y_i}(\omega)$, can be expressed as [48],

$$\Phi_{y_i}(\omega) = \frac{\gamma_i^2}{\omega^2} \tag{91}$$

Let us consider a r.v. Z which is obtained by magnitude scaling of a r.v. Y i.e., $Z = \delta Y$, the characteristic function of Z, $\Phi_z(\omega)$, in terms of $\Phi_y(\omega)$ can be expressed as [48],

$$\Phi_z(\omega) = \Phi_z(\delta\omega) \tag{92}$$

Therefore, the characteristic function of r.v. S_i^{J-1},$\Phi_{s_i^{J-1}}(\omega)$, can be expressed as

$$\Phi_{s_i^{J-1}}(\omega) = \prod_{k=1}^{N_1} \Phi_{y_k}\left(\delta_{k-2i}\omega\right) \tag{93}$$

$$= \prod_{k=1}^{N_1} \frac{\gamma_i^2}{\left(\gamma_i^2 + \left(\delta_{k-2i}\omega\right)^2\right)} \tag{94}$$

$$= \prod_{k=1}^{N_1} \frac{\acute{\gamma}_i^2}{(\acute{\gamma}_i^2 + \omega^2)} \tag{95}$$

where $\acute{\gamma}_i = \gamma_i/\gamma_{k-2i}$ and N_1 is the length of the wavelet filter.

In order to determine the pdf of wavelet coefficients S_i^{J-1}, $f_{s_i^{J-1}}(\tau)$ characteristic function $\Phi_{s_i^{J-1}}(\omega)$ (given by Eq. (95)) is used. The pdf of a r.v. can be determined either using the uniqueness theorem or the convolution theorem [48]. The pdf of wavelet coefficients S_i^{J-1}, $f_{s_i^{J-1}}(\tau)$ using $\Phi_{s_i^{J-1}}(\omega)$ based on the convolution theorem can be expressed,

$$f_{s_i^{J-1}}(\tau) = \frac{\acute{\gamma}_i}{2} e^{-\acute{\gamma}_i|t|} \left(\sum_{k=0}^{N_1-1} c_i^k \acute{\gamma}_i^k t^k \right) \tag{96}$$

where $c_k \in \mathcal{R}$ is a real constant.

For different values of N_1, the polynomial coefficients, c_k, are given as:

$N_1 = 2, c_0 = c_1 = \frac{1}{2}$, and $N_1 = 3, c_0 = c_1 = \frac{3}{8}$, and $c_1 = \frac{1}{8}$ and so on.

According to the Eq. (95) and (96), as the pdf of wavelet coefficients at j_{th} level, $f_{s_i^{j-1}}(\tau)$, is a obtained by convolving the pdf of r.v. Y_i, therefore, based on the CLT, the pdf of the subband coefficients move to words Gaussianity as value of N_1 increases or in other words, pdf of wavelet coefficients at coarser level is closer to the Gaussianity than higher level coefficients. This is because at coarser level, for each wavelet coefficient more audio samples contribute in the weighted-sum equation (given by Eq. (89)) than higher level coefficients.

In order to provide evidence in support of this model, a 4–level DWT decomposition of an arbitrary frame of the music clip *I Want It That Way*$\cdots$ by *Backstreet Boys*, using '*Daubechies*-8' decomposition filter, is given in Fig. 13. The pdf (based on histogram approximation) of corresponding wavelet coefficients at different levels is plotted in Fig. 13. This is clear from Fig. 13 that the higher level, wavelet coefficients exhibit non-Gaussian distribution and distribution moves towards Gaussianity for coarser coefficients due to longer weighted-sum effect at the coarser level.

Therefore, the pdf of each subband coefficient (at higher level) of the host signal, S_i, can be approximated by Laplacian distribution, which is given as,

$$f_s(\tau) = \frac{\beta}{2} e^{-\beta|\tau|} : |\tau| < \infty \tag{97}$$

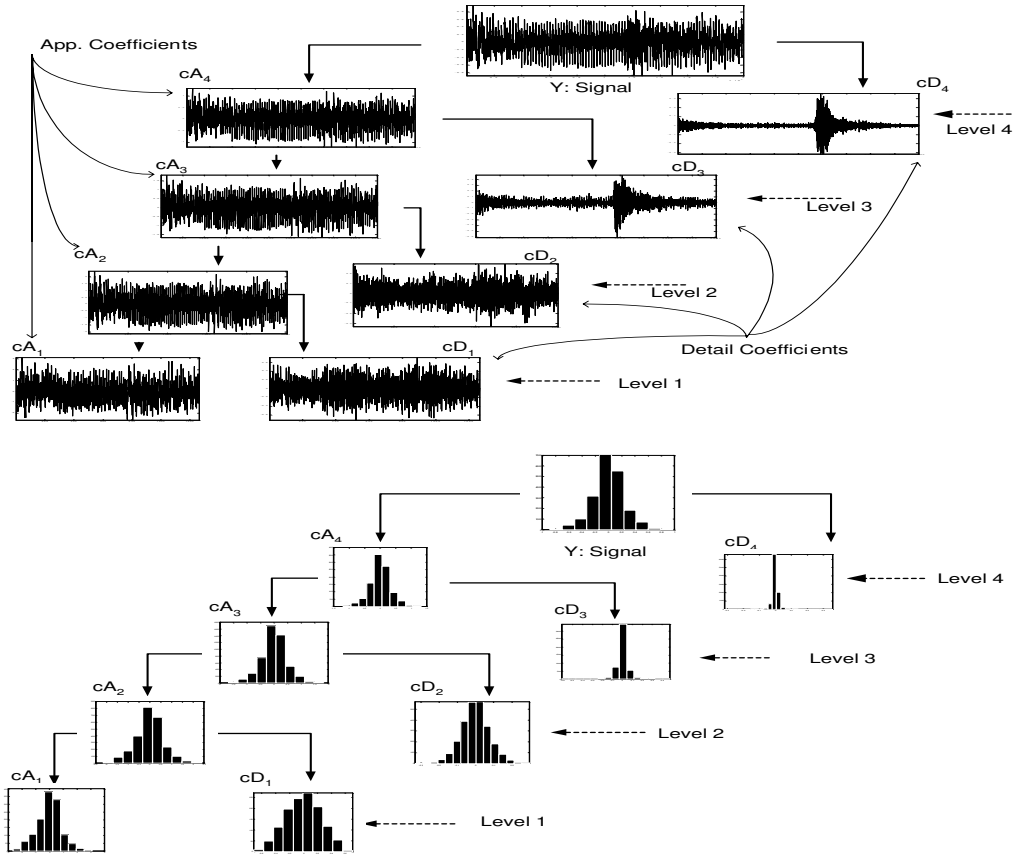

Figure 13: Plots of empirical distribution based histogram approximation of detailed and approximate coefficients at each level of 4-Level wavelet decomposition of an audio signal, $\mathbf{y}$

where $\beta = \frac{\sqrt{2}}{\sigma_s^2}$