ON THE DETECTION OF MULTIPLICATIVE WATERMARKS FOR SPEECH SIGNALS IN THE WAVELET AND DCT DOMAINS

Ramin Eslami, J. R. Deller, Jr., and Hayder Radha
Email: {eslamira , deller , radha} @ egr.msu.edu

ECE Department / 2120 EB, Michigan State University, East Lansing, MI 48824, USA

ABSTRACT

Blind multiplicative watermarking schemes for speech signals using wavelets and discrete cosine transform are presented. Watermarked signals are modeled using a *generalized Gaussian distribution* (GGD) and Cauchy probability model. Detectors are developed employing *generalized likelihood ratio test* (GLRT) and *locally most powerful* (LMP) approach. The LMP scheme is used for the Cauchy distribution, while the GLRT estimates the gain factor as an unknown parameter in the GGD model. The detectors are tested using Monte Carlo simulation and results show the superiority of the proposed LMP/Cauchy detector in some experiments.

1. INTRODUCTION

Spread-spectrum approaches are among the traditional methods for efficient watermarking [5][11]. In these techniques, a pseudo-random noise sequence is adopted as the watermark and embedded into the transform domain [normally the *discrete cosine transform* (DCT) or *wavelet transform* (WT)] of the multimedia content. Embedding the watermark in the perceptually most significant transform coefficient(s) of the media, one would expect a more robust watermark against possible attacks [5][11].

Watermarks are ordinarily embedded using linear combinations in the relevant domain. In this paper, multiplicative watermarks are employed to enhance the complexity of the embedding operation, thereby rendering it more difficult to detect, remove, or destroy the watermark [9]. In addition to increased robustness to attack, multiplicative watermarks have been noted to possess another important feature for speech, the signal of interest in the present work. In fact, multiplicative embedding rule is in accordance with the Weber's law and thus the *human visual system* (HVS) relatively insensitive to this type of watermark [9]. Remarkably, we also found that embedding multiplicative watermark in a (speech) signal would degrade the quality of that signal less than an additive watermark using the same gain factor.

Blind watermark detection has been an area of active research in recent years. Assuming a *generalized Gaussian distribution* (GGD) model for the DCT coefficients of images, the authors in [7] propose a *likelihood ratio test* (LRT) watermark detector in which an additive rule is used and the gain factor is assumed known. Although the LRT detection provides optimality in the Neyman-Pearson sense, its existence is subject to the availability of the gain factor γ. The gain factor is generally unknown. Note that since the watermark is under attack, one cannot embed the gain factor in the watermark. Further, it is signal dependent and hence it is not possible to keep it in the public key. It follows that the optimal LRT detector cannot be realized. Another example of watermark detection is the method proposed by Cheng and Huang [4], where a *locally most powerful* (LMP) detection of multiplicative watermarks with a GGD model employed. Finally, Briassouli and Strintzis [3] used Cauchy and Gaussian-tailed zero memory nonlinearities to better capture the heavy tail of the DCT coefficients. Then they applied the proposed model for additive watermark detection in the DCT domain.

In this work, we develop a *generalized likelihood ratio test* (GLRT) as well as LMP tests for the detection of multiplicative watermarks in the WT and DCT domains. We model the distribution of watermarked coefficients in these domains with a GGD model and also a Cauchy *probability distribution function* (PDF). The proposed watermarking strategies are then tested for robustness to certain attacks.

2. SPEECH WATERMARKING AND PROBABILITY MODELS

2.1. Spread-Spectrum Watermarking

Spread-spectrum watermarking in the DCT and WT domains has shown potential for image watermarking [5][11]. In this work, we apply this watermarking technique to speech signals.

Let the watermark $w_i\ (0 \le i < N)$ be a realization of white Gaussian noise with PDF $\mathcal{N}(0,1)$. Let $s_i\ (0 \le i < N_s)$ be the original signal and $y_i\ (0 \le i < N_s)$ denote the transform (DCT or WT) coefficients of the signal. In order to embed the watermark in the perceptually most significant coefficients of the signal, the N transform coefficients with the largest magnitudes [$x_i\ (0 \le i < N)$] are selected for embedding. In the DCT domain this translates to the selection of high frequency coefficients; however, the high-magnitude (or *significant*) WT coefficients represent the transitions in the signal which include a significant amount of the information content of speech [6]. In fact, this selection is in accordance with the *human auditory system* (HAS) perceptual model. Accordingly, the watermarked signal $z_i\ (0 \le i < N)$ can be expressed as

$$z_i = x_i(1+\gamma w_i) \quad \text{for} \quad 0 \le i < N , \tag{1}$$

where we employ a single gain factor γ. Sections 2.2 and 2.3 discuss the impact of the signal PDFs on the selection of significant transform coefficients.

This work was supported by the National Science Foundation under Cooperative Agreement IIS-9817485. Conclusions and recommendations expressed in this material are those of the authors and not necessarily those of the NSF.

2.2. Generalized Gaussian Distribution Model

The GGD is an appropriate PDF for modeling WT and DCT coefficients of images [2][7]. We justify below that the GGD can also be used to effectively model the significant transform coefficients of speech. To illustrate the distribution of a speech signal ($s1$) in the WT and DCT domains, normalized histograms of the significant coefficients are shown in Fig. 1. Here, $N_{s1} = 46{,}000$ is the signal size and $N = 10{,}000$ is the number of significant coefficients. The histograms are clearly non-Gaussian whereas a GGD model would be a proper choice. Note that the small gap in the middle of histograms is due to the exclusive selection of high-amplitude coefficients. Simulation results confirm that such "gaps" have little impact on the detector performance.

A GGD with zero mean is defined as $p_X(x) = A e^{-|\beta x|^c}$ ([2]) where $\beta = (1/\sigma)\left(\Gamma(3/c)/\Gamma(1/c)\right)^{1/2}$, $A = \beta c \big/ \left(2\Gamma(1/c)\right)$, and Γ denotes the gamma function. Here, c is the shape parameter, which is equal to two for a Gaussian distribution and one for the Laplacian PDF. In the present work, the parameters are estimated using the method proposed in [2].

For the above example, c equals 0.50 for the WT and 0.90 for the DCT. The fitted GGDs are illustrated in Fig. 1. Since a low gain factor is used to assure inaudibility of the watermark, the histograms remain practically unchanged, so that the GGD remains an appropriate model.

To obtain the probability of the watermarked signal z (in the transform domain), we assume that the watermarked coefficients $z_i \ (0 \le i < N)$ are *independent and identically distributed* (i.i.d.). Therefore, from (1) we can compute $p_Z(z)$ as

$$p_Z(z) = \prod_{i=1}^{N} \left(A \big/ |1+\gamma w_i|\right) e^{-\sum_{i=1}^{N} |\beta z_i / (1+\gamma w_i)|^c} , \qquad (2)$$

where, $z = (z_1, \ldots, z_n)$ is the watermarked signal and the gain factor γ is the unknown parameter.

2.3. Cauchy Distribution Model

Unlike the Gaussian PDF and GGD, the Cauchy probability model does not taper quickly making it a proper model for heavy-tailed distributions. The PDF of a zero-mean Cauchy distribution is $p_X(x) = \alpha \big/ \left(\pi(x^2 + \alpha^2)\right)$. We can estimate the parameter α using a *maximum likelihood* (ML) approach. To do so, we consider the $x_i \ (0 \le i < N)$ to be i.i.d. random variables, where each x_i has a Cauchy PDF. Thus, $p_X(x) = \prod_{i=1}^{N} (\alpha/\pi)\left(1 \big/ (x_i^2 + \alpha^2)\right)$, where $x = (x_1, \ldots, x_n)$.

To estimate α, we compute $d \ln p_X(x) / d\alpha = 0$, which leads to $J = N/\alpha - \sum_{i=1}^{N} 2\alpha \big/ (\alpha^2 + x_i^2) = 0$. Using the Newton's algorithm, α can be obtained via the recursion $\alpha^{[n+1]} = \alpha^{[n]} - J(\alpha^{[n]}) \big/ J'(\alpha^{[n]})$, where $J' = dJ/d\alpha$.

For the signal $s1$ discussed in Section 2.2, the estimated α is equal to 0.058 and 0.097 for the significant WT and DCT coefficients, respectively. Fig. 1 shows with dashed lines the Cauchy PDFs fitted to the histograms.

Using the i.i.d. assumption for the watermarked signal z (in the WT or DCT domain) as in the case of the GGD model, we can express $p_Z(z)$ as

$$p_Z(z) = \prod_{i=1}^{N} (\alpha/\pi)\left(|1+\gamma w_i| \big/ [z_i^2 + \alpha^2 (1+\gamma w_i)^2]\right) . \qquad (3)$$

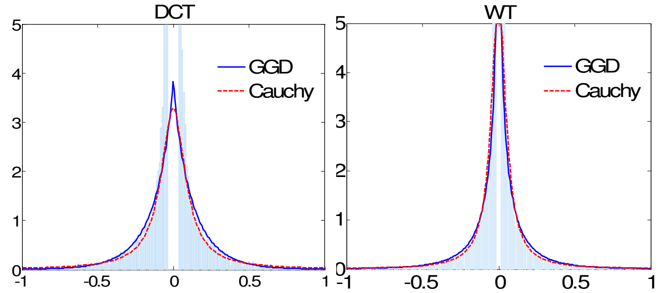

Fig. 1. Normalized histograms of the significant transform coefficients of the speech signal $s1$ as well as the fitted GGD and Cauchy distribution.

3. GENERALIZED LIKELIHOOD RATIO TEST

In GGD or Cauchy distribution an unknown parameter, i.e., the gain factor γ, renders it impossible to take advantage of the *likelihood ratio test* (LRT), which is an optimal detector in the Neyman-Pearson sense. For a given probability of false alarm, P_f, the probability of detection, P_d, is maximized. Nevertheless, one can employ ML estimation of γ and employ the GLRT [8].

3.1. Maximum Likelihood Estimation of the Gain Factor

In the case of Cauchy PDF, after taking the derivative of logarithm of $p_Z(z)$ in (3) and setting it to zero, we obtain

$$\partial \ln p_Z(z) / \partial\gamma = \sum_{i=1}^{N} w_i (z_i^2 - \alpha^2 \varphi_i^2) \big/ \left(\varphi_i (z_i^2 + \alpha^2 \varphi_i^2)\right) = 0 ,$$

where $\varphi_i = 1 + \gamma w_i$. A closed-form solution for γ cannot be derived. Moreover, numerical methods return multiple solutions. Thus, we merely use the LMP test for the Cauchy distribution.

Here we obtain an ML estimation of γ for use with the GGD in (2). For the GGD model, the optimization equation is

$$\partial \ln p_Z(z) / \partial\gamma = \sum_{i=1}^{N} \left(c w_i |\beta z_i| \varphi_i^{-(c+1)} - w_i \varphi_i^{-1}\right) = 0 .$$

Again, a closed-form solution of γ is not achievable. However, since the gain factor is a small number (usually less than 0.1), we have $|\gamma w_i| < 1$ and therefore, one could take advantage of binomial identity to approximate $\varphi_i^{-(c+1)}$ and φ_i^{-1} to within two orders to obtain the following quadratic equation, $A\hat{\gamma}^2 + B\hat{\gamma} + C = 0$, where

$$A = (1/2)c(c+1)(c+2) \sum_{i=1}^{N} w_i^3 |\beta z_i|^c ,$$

$$B = \sum_{i=1}^{N} w_i^2 \left(1 - c(c+1)|\beta z_i|^c\right), \text{ and } C = \sum_{i=1}^{N} c w_i |\beta z_i|^c .$$

The above equation provides two solutions for $\hat{\gamma}$. The acceptable solution satisfies $|\hat{\gamma}| < 1$. In the next section, the GLRT using the estimated $\hat{\gamma}$ is developed.

3.2. GLRT for the GGD probability model

To detect the watermark $w = (w_1, \ldots, w_n)$, we consider the hypotheses pair $H_0: \gamma = 0$ vs. $H_1: \gamma \ne 0$. Now we form the GLRT using $p_Z(z; \gamma = \hat{\gamma}; H_1)$ and $p_Z(z; \gamma = 0; H_0)$, which becomes

$$L_G(z) = p_Z(z; H_1) \big/ p_Z(z; H_0) = \prod_{i=1}^{N} \left|\hat{\varphi}_i^{-1}\right| e^{-\sum_{i=1}^{N} \left(|\beta z_i / \hat{\varphi}_i|^c - |\beta z_i|^c\right)}$$

or

$$T_{GLRT}^{(GGD)}(z) = \ln L_G(z) = \sum_{i=1}^{N} \left(\ln\left|\hat{\varphi}_i^{-1}\right| - \left|\beta z_i / \hat{\varphi}_i\right|^c - \left|\beta z_i\right|^c\right),$$

where $\hat{\varphi}_i = 1 + \hat{\gamma} w_i$. The GLRT detection scheme is similar to the LRT with the exception that estimates of the unknown parameters are used. The LMP detector, however, does not require parameter estimation as explained next.

4. LOCALLY MOST POWERFULL TEST

A LMP detector in conjunction with the GGD model is proposed in [4] for multiplicative watermark detection. In this work, however, we propose a LMP detector using a Cauchy PDF for modeling the transform coefficients. Below, a brief description of the LMP test is provided.

If $0<\gamma\ll 1$, one can approximate $p_Z(z;\gamma)$ as

$$\ln p_Z(z;\gamma)=\ln p_Z(z;0)+\partial \ln p_Z(z;\gamma)/\partial\gamma\big|_{\gamma=0}\,\gamma\,,$$

using a first-order Taylor expansion. Recall that the LRT is expressed as $T_{LRT}(z)=\ln\left(p_Z(z;\gamma)/p_Z(z;0)\right)>\theta$, where θ is the threshold. Hence, $T_{LRT}(z)=\partial \ln p_Z(z;\gamma)/\partial\gamma\big|_{\gamma=0}\,\gamma>\theta$ or $\partial \ln p_Z(z;\gamma)/\partial\gamma\big|_{\gamma=0}>\theta/\gamma=\tilde{\theta}$. The scaled LMP test is [8]

$$T_{LMP}(z)=\partial \ln p(z,\gamma)/\partial\gamma\big|_{\gamma=0}\sqrt{\mathrm{I}^{-1}(\gamma)\big|_{\gamma=0}}\,, \quad (4)$$

where $\mathrm{I}(\gamma)=-E\left[\partial^2 p(z;\gamma)/\partial\gamma^2\right]$ is the Fisher information. Here, we have assumed that the hypotheses imply a one-sided test

$$H_0:\ \gamma=0 \ \text{ vs. } \ H_1:\ \gamma>0\,.$$

Since T_{LMP} is derived from the optimal detector, T_{LRT}, its performance is optimal for small values of γ. For large data records (when N is large), the LMP statistic in (4) has a Gaussian PDF under each hypothesis as [8],

$$T_{LMP}(z)\xrightarrow[N\text{ large}]{}\begin{cases}\mathcal{N}(0,1), & \text{under } H_0\\ \mathcal{N}(\gamma\sqrt{\mathrm{I}(\gamma)\big|_{\gamma=0}},\ 1), & \text{under } H_1\end{cases}$$

Therefore, the LMP test comprises a Gauss-Gauss detector with the *deflection coefficient* equal to $d_{LMP}=\gamma\sqrt{\mathrm{I}(\gamma)\big|_{\gamma=0}}$. Accordingly, the theoretical asymptotic *receiver operating characteristic* (ROC) will be

$$P_d=Q\left(Q^{-1}(P_f)-d_{LMP}\right), \quad (5)$$

where $Q(\theta)=(1/2\pi)\int_{\theta}^{\infty}e^{-(1/2)t^2}\,dt$.

4.1. LMP Detector Using the GGD

We can construct the LMP detector by applying (4) to the GGD model (2). After some calculations we obtain

$$T_{LMP}^{(GGD)}(z)=\frac{\sum_{i=1}^{N}c w_i\left|\beta z_i\right|^c}{\sqrt{c(c+1)\sum_{i=1}^{N}\left|\beta z_i\right|^c-N}}\,,$$

where the watermark w_i is assumed to have a normal PDF $\mathcal{N}(0,1)$. Also note that

$$E\left[w_i^2\left|\beta z_i\right|^c\right]=E[w_i^2]E\left[\left|\beta z_i\right|^c\right]=E\left[\left|\beta z_i\right|^c\right]\simeq\left|\beta z_i\right|^c\,.$$

4.2. LMP Detector Using the Cauchy PDF

Because of the Cauchy PDFs appropriateness for heavy-tailed distributions, Briassouli and Strintzis [3] developed an LMP/Cauchy detector for additive watermark detection, and showed its efficiency for detecting the watermarks in the DCT domain. Here we develop a LMP detector for multiplicative watermarks. Taking the derivative of logarithm of (3) and evaluating the result at $\gamma=0$ yields

$$\partial \ln p_Z(z)/\partial\gamma\big|_{\gamma=0}=\sum_{i=1}^{N}w_i(z_i^2-\alpha^2)/(z_i^2+\alpha^2)\,.$$

Further, the Fisher information at $\gamma=0$ is computed as

$$\mathrm{I}(\gamma)\big|_{\gamma=0}=N+2\alpha^2\sum_{i=1}^{N}E\left[(z_i^2-\alpha^2)/(z_i^2+\alpha^2)^2\right],$$

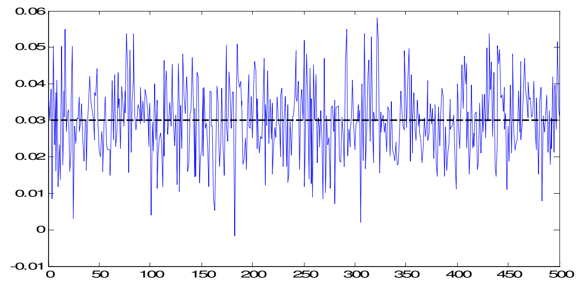

Fig. 2. Estimated values of the gain factor using the DCT coefficients of the watermarked speech signal $s1$.

where we again approximate the expectation by its experimental value. Consequently, the LMP test statistic is

$$T_{LMP}^{(Cauchy)}(z)=\frac{\sum_{i=1}^{N}w_i(z_i^2-\alpha^2)/(z_i^2+\alpha^2)}{\sqrt{N+2\alpha^2\sum_{i=1}^{N}E\left[(z_i^2-\alpha^2)/(z_i^2+\alpha^2)^2\right]}}\,.$$

5. EXPERIMENTAL RESULTS

The speech signal used in the experiments, designated $s1$ above, is an utterance of the sentence “She had your dark suit in greasy wash water all year,” spoken by a female in the TIMIT database [10], and the watermark is a realization of zero-mean white Gaussian noise with $\sigma=1$. To achieve inaudibility of the watermark, we set $\gamma=0.08$ for the WT watermarking and $\gamma=0.03$ for the DCT. Remarkably, even with this lower gain factor in the DCT domain, a background noise is heard upon high volume playback.

The developed detectors were tested via Monte Carlo simulation. To do so, we generated 500 realizations of watermarks. Then at each time we embedded a watermark, then attempted detection of each of the 500 watermarks in the stegosignal. Thus, a total of 250,000 detections were executed by each detector to generate an ROC curve. To implement the GLRT detector using the GGD, the gain factor had to be estimated at each run. Fig. 2 shows the estimated values of γ using the DCT significant coefficients. The average of these values is 0.03, which is the same as the true value in this case.

Detectors performance was first assessed without attacks on the stegosignals. In the WT case, perfect detection resulted over all experiments. Accordingly, the gain factor was lowered to $\gamma=0.04$. We also obtained the asymptotic values of the LMP detectors using (5). Moreover, we applied the optimal LRT schemes for the sake of comparison. Fig. 3 demonstrates the resultant ROC curves in the WT and DCT domains. As seen, the LMP detector with a Cauchy PDF yields the best results in both domains. Furthermore, the LMP detection strategies nearly provide identical performance to their related optimal LRT detectors. The GLRT technique yields suboptimal results. In addition, the asymptotic theoretical ROC curves obtained from (5) for the LMP detection schemes are slightly different from the empirical ones. This is likely due to the fact that the model PDFs do not exactly fit the data.

To evaluate the robustness of the proposed watermarking schemes, several severe attacks were applied to the stegosignal. In the cropping attack, only 50% of the signal was retained in the middle of the waveform. A downsampling attack removed seven

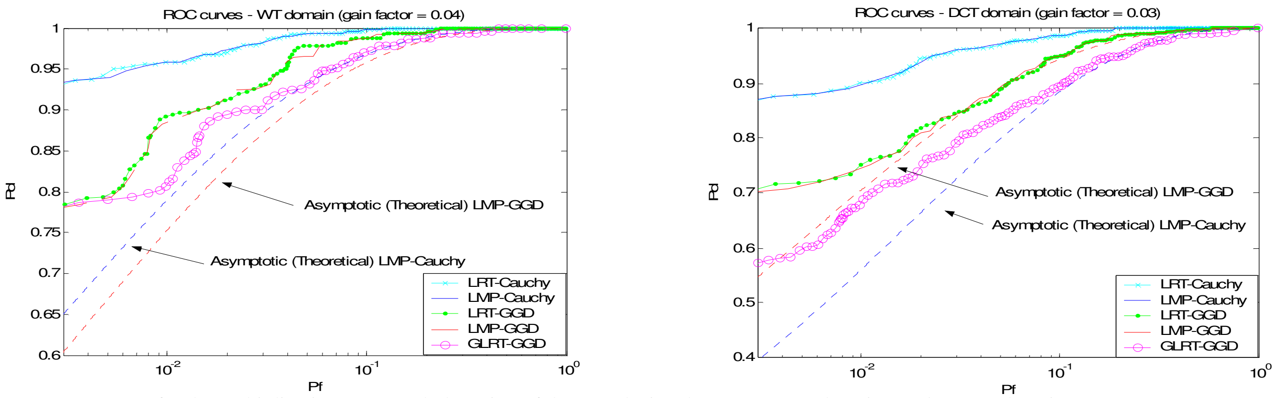

Fig. 3. ROC curves for the multiplicative watermark detection of the speech signal $s1$. *Left:* WT domain. *Right:* DCT domain.

of eight samples, and a filtering attack lowpass-filtered the watermarked signal to remove energy above 3.5 kHz. Only the LMP detectors were employed due to their optimal performance in the "clean" experiments. Fig. 4 shows the resulting ROC curves for both WT and DCT embedding. Despite the superiority of the LMP/Cauchy detector with no attack, this detector performs more poorly than the LMP detector using the GGD model in the wavelet domain. In the case of the DCT watermarking, however, the Cauchy PDF appears to be a better model for the DCT coefficients in most cases. This is because the DCT coefficients distribution has heavy tails which are better modeled by the Cauchy PDF. Remarkably, the WT watermarking strategy is more robust to filtering attack than the DCT scheme. Since the watermark is embedded in the high frequency components of signal in the DCT approach, however, DCT embedding is not robust to lowpass filtering attacks.

6. CONCLUSION

Watermarking strategies for speech signals using the WT and DCT significant coefficients have been presented. The WT approach is more appropriate for embedding since the significant coefficients of this scheme represent the transitions to which the human auditory system is less sensitive. A method for estimation of the gain factor was incorporated into a GLRT detection strategy. A LMP detector in conjunction with the Cauchy PDF has also been developed and found to be effective when no attack is involved. Further, the proposed LMP/Cauchy detector is superior to the LMP/GGD scheme for DCT watermarking.

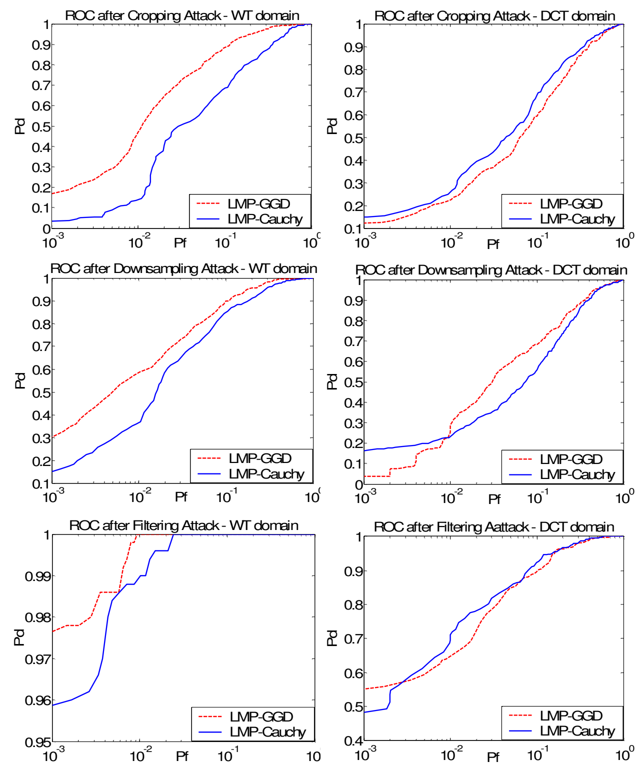

Fig. 4. ROC curves obtained from the attacked watermarked signal $s1$. *Top:* Cropping to 50% of the original size. *Middle:* Downsampling by 8. *Bottom:* Lowpass filtering with cutoff frequency of 3.5 kHz.

7. REFERENCES

[1] M. Barni, F. Bartolini, A. D. Rosa, and A. Piva, "A new decoder for the optimum recovery of the nonadditive watermarks," *IEEE Trans. Image Processing,* vol. 10, pp. 755-766, May 2001.

[2] K. A. Birney and T. R. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Trans. Image Processing,* vol. 4, pp. 186-193, Feb. 1995.

[3] A. Briassouli and M. G. Strintzis, "Locally optimum nonlinearities for DCT watermark detection," *IEEE Trans. Image Processing,* vol. 13, no. 12, pp. 1604-1617, Dec. 2004.

[4] Q. Cheng and T. S. Huang, "Robust optimum detection of transform domain multiplicative watermarks," *IEEE Trans. Signal Processing,* vol. 51, no. 4, pp. 906-924, Apr. 2003.

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

[6] J.R. Deller, Jr., J.H.L. Hansen, and J.G. Proakis, *Discrete-Time Processing of Speech Signals.* (2d ed.), IEEE Press / Wiley, 2000.

[7] J. R. Hernandez, M. Amado, and F. Perez-Gonzalez, "DCT-domain watermarking techniques for still images: detector performance analysis and a new structure," *IEEE Trans. Image Processing,* vol. 9, pp. 55-68, Jan. 2000.

[8] S. M. Kay, *Fundamentals of Statistical Signal Processing, vol. II: Detection Theory.* NJ, Prentice Hall, 1998.

[9] G. Langelaar, I Setyawan, and R. Lagendijk, "Watermarking digital image and video data," *IEEE Signal Processing Magazine,* pp. 20-46, Sep. 2000.

[10] National Institute of Standards and Technology, "Getting Started with the DARPA TIMIT CD-ROM," Gaithersburg MD, 1988.

[11] C. Podilchuk and W. Zeng, "Image adaptive watermarking using visual models," *IEEE JSAC,* vol. 16, pp. 525-539, May 1998.