DETECTOR DESIGN FOR PARAMETRIC SPEECH WATERMARKING

A. Gurijala and J. R. Deller, Jr.

Michigan State University
Dept. of Electrical & Computer Engineering / 2120 EB
East Lansing, MI 48824 USA

ABSTRACT

Parametric watermarking is effected by modifying the linear predictor coefficients of speech. In this work, the parameter noise is analyzed when watermarked speech is subjected to additive white and colored noise in the time domain. The paper presents two detection techniques for parametric watermarking. The first approach uses the Neyman-Pearson criterion to solve a binary decision problem. In the second approach, discriminant functions based on the minimum-error-rate criterion are used to determine which one of the many watermarks was embedded or if no watermark is present. Experiments with speech data are used to determine the false-alarm and missed detection rates.

1. INTRODUCTION

Digital watermarking has emerged as a new technology for the protection of copyrighted material. Digital watermarking is the process of embedding data (*watermark*) imperceptibly into the host signal (*coversignal*), resulting in the *stegosignal*. When copyright questions arise, the watermark is recovered from the stegosignal as evidence of title. The design of a watermarking strategy involves the balancing of two principal criteria. First, embedded watermarks must be imperceptible to the listener. Second, watermarks must be robust. That is, they must be able to survive *attacks* - those deliberately designed to destroy them, as well as distortions inadvertently imposed upon the watermarks by technical or systemic processes.

Parametric watermarking [1] is based on manipulation of linear predictor parameter values [2] of speech signals. This paper deals with watermark detector design for parametric watermarking. A common approach to watermark detection involves the hypotheses, $H_0 : I_R = I$ and $H_1 : I_R = I + W$, where I_R is the received signal, I is the original signal and W is the watermark signal [3], [4]. A Bayesian or Neyman-Pearson approach is followed in deriving the detection thresholds. For image watermarking, the image DCT coefficients are generally modeled as generalized Gaussian in distribution [3]. Most of these approaches do not consider the effect of noise while deriving the detection threshold. Several watermark detectors are based on correlation detection [5], [6]. That is, the correlation between the original and recovered watermarks or the correlation between the original watermark and recovered signal is compared against a threshold. Correlation detectors are optimal when the watermark and noise are jointly Gaussian, or in case of blind detectors the watermarked signal and noise should be jointly Gaussian. For example, the detector presented in [7, Ch. 6] assumes that the detector output for each bit is Gaussian distributed. This is true for watermark patterns that are white, which is not the case in parametric watermarking. In this work, noise in the parameter domain is analyzed when the stegosignal is distorted by additive white or colored Gaussian noise in the time domain. Watermark detection in the parameter domain is converted into a binary or multiple decision problem in the presence of additive noise.

2. BACKGROUND

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

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

The "true" model is determined by standard LP analysis of a frame selected for watermarking [2]. The sequence $\{\xi_n\}$ is the prediction residual associated with the estimated model. The stegosignal is constructed using the FIR filter model

$$\tilde{y}_n = \sum_{i=1}^{M} \tilde{a}_i y_{n-i} + \xi_n \tag{2}$$

where $\{\tilde{a}_i\}$ represents a deliberately perturbed version of the "true" set $\{a_i\}$. The algorithmic steps for watermark embedding and recovery appear in Tables 1 and 2, respectively.

An important step in the recovery process, is the least squares estimation of modified watermark coefficients. In the stegosignal, the watermark information is spread out, while during recovery the watermark information is concentrated in a few coefficients $\{\omega_i\}_{i=1}^{M}$ derived from an estimate of the modified LP coefficients. Parametric watermarking involves *informed detection*, and the coversignal is required for watermark recovery. CWR_{seg} is used to measure stegosignal fidelity and is defined as,

$$\text{CWR}_{\text{seg}} = \frac{1}{K}\sum_{j=1}^{K} 10\log_{10}\left[\sum_{l=k_j-L+1}^{k_j} \frac{y_l^2}{[\tilde{y}_l - y_l]^2}\right], \tag{3}$$

where, $k_1, k_2, ..., k_K$ are the end-times for the K frames, each of which is length L. The CWR_{seg} assigns equal weight to the loud and soft portions of speech. For computing CWR_{seg}, speech frames of 15 ms duration were used. A simple way to control the fidelity of the stegosignal is to scale the watermark vector, ω, by a constant, say κ, before adding it to the original LP parameters.

This work was supported by the National Science Foundation of the United States under Cooperative Agreement No. IIS-9817485. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

0-7803-9332-5/05/$20.00 ©2005 IEEE

Table 1. WATERMARK EMBEDDING ALGORITHM

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

1 Using the "autocorrelation method" (e.g., [2, Ch. 5]), derive a set of LP coefficients of order M, say $\{a_i\}_{i=1}^{M}$, for the given frame. Use the LP parameters in an *inverse filter* configuration (e.g. [2, Ch. 5]) to obtain the prediction residual on the frame, $\left\{\xi_n = y_n - \sum_{i=1}^{M} a_i y_{n-i}\right\}_{n=n_k}^{n'_k}$.

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

3 Use the modified LP parameters as a (suboptimal) predictor of the original sequence, adding the residual obtained in Step 2 above at each n, to resynthesize the speech over the frame, $\left\{\tilde{y}_n = \sum_{i=1}^{M} \tilde{a}_i y_{n-i} + \xi_n\right\}_{n=n_k}^{n'_k}$. The sequence $\{\tilde{y}_n\}_{n_k}^{n'_k}$ is the k^{th}frame of the watermarked speech (stegosignal).

Next k.

Table 2. WATERMARK RECOVERY ALGORITHM

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

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

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

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

Next k.

3. WATERMARK DETECTION IN PARAMETER DOMAIN

The watermarks are comprised of non-binary orthogonal vectors of length eight. Each of the eight orthogonal vectors (ω^k for $k = 1, \cdots, 8$) can be interpreted as symbols from an alphabet of size eight. The watermark may be composed of many such symbols. Each orthogonal watermark vector (symbol) of length eight, is embedded into 0.125 seconds of speech, sampled at 16 kHz. The watermark vector is added to the coefficients of an eighth order linear predictor model. The length of the watermark vector (and hence the predictor model order) and the duration of speech frame can be selected quite arbitrarily, subject to certain constraints on stegosignal fidelity. These constraints include an upper limit on predictor model order, and a need to use FIR models of small order for very short speech frames (˜500 samples).

Noise in the LP domain, caused by stegosignal exposure to additive noise, can be modeled as having a Gaussian distribution. Figure 1(a) shows a typical noise distribution in the LP domain when white noise (SNR 15 dB) is added to the stegosignal. This noise distribution (Fig. 1(a)) was obtained by conducting 1000 experiments, involving a stegosignal of 1 s duration watermarked at CWR$_{\text{seg}}$ of 7 dB using a watermark message consisting of eight orthogonal vectors, each vector embedded into 0.125 seconds or 2000 samples of speech. The coversignal was the sentence from TIMIT database [8], "She had your dark suit in greasy wash water all year," sampled at 16 kHz. When white Gaussian noise was added to the stegosignal, the noise samples affecting a particular watermark coefficient, in the parameter domain were uncorrelated and could be approximated as independent and identically distributed (i.i.d) Gaussian noise. The LP noise affecting a watermark coefficient was uncorrelated with the LP noise affecting a different watermark coefficient. The noise samples were also uncorrelated with the corresponding LP coefficients. When the stegosignal is distorted by white Gaussian noise, the parameter noise is of very low power and asymptotically tends to zero. Also, the noise generated using the "randn" function in matlab, is not ideal white noise.

The parametric noise distribution when the stegosignal was affected by colored noise is similar to that shown in Fig. 1(b). Colored noise was generated by filtering a white noise process using a 11^{th}order FIR lowpass filter with a cut-off frequency of 6400 Hz. The LP noise affecting any given watermark coefficient was found to be Gaussian and i.i.d. in nature. However, a realization of noise affecting all the watermark coefficients was found to be correlated with the original LP coefficients. A possible solution to this problem is to normalize the watermark coefficients before adding them to the original LP coefficients. That is, instead of directly adding the watermark vector to the original LP coefficients ($\tilde{\mathbf{a}} = \mathbf{a} + \omega$, where ω is any of the eight orthogonal vectors ω^k), we obtain the modified LP coefficients as,

$$\tilde{a}_i = a_i + \omega_i |a_i|. \tag{4}$$

From the estimate of the modified LP coefficients, the watermark vector vector is obtained as,

$$\hat{\omega}_i = \frac{\hat{\tilde{a}}_i - a_i}{|a_i|} \tag{5}$$

with $\hat{\tilde{\mathbf{a}}} = \{a_i\}_{i=1}^{M}$, as defined in Table 2. However, when $|a_i| \ll$

Table 3. EFFECT OF SELECTIVE NORMALIZATION

Noise	**SNR (dB)**	**Normalization**	μ	σ^2	$c_{\mathbf{ra}}(0)$
White	10	no	2.849×10^{-4}	0.0517	-0.0059
White	10	complete	-0.0152	4.6477	-0.0051
White	10	selective	-6.2×10^{-5}	0.1099	7.6445×10^{-4}
Colored	15	no	2.3438×10^{-4}	0.0049	0.0328
Colored	15	complete	0.0139	0.7518	-0.0094
Colored	15	selective	-1.162×10^{-4}	0.0071	0.0023

1, the recovery of watermark coefficients magnifies the noise variance in the LP domain. To avoid this, watermark coefficients are normalized before embedding only if $|a_i| \geq 1$. For the experiments presented in this paper, watermark embedding and recovery involves this "selective normalization." Step 3 of Table 1 is carried out using the following rule in the algorithm implemented here:

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

The final step in the recovery algorithm (Table 2) involves the following equation:

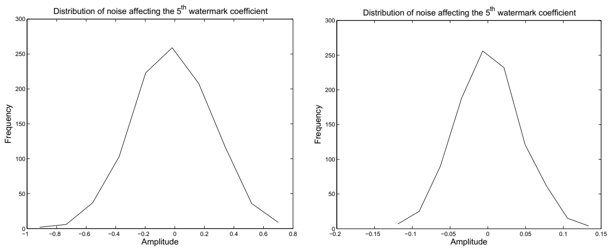
$$\hat{\omega}_i = \begin{cases} \frac{\hat{\tilde{a}}_i - a_i}{|a_i|}, & \text{if } |a_i| \geq 1 \\ \hat{\tilde{a}}_i - a_i, & \text{otherwise} \end{cases}$$

It is observed from Table 3 that selective normalization of watermark coefficients significantly reduces the correlation between noise in LP domain and the LP coefficients, especially when the stegosignal is subjected to colored noise in the time domain. As the noise variance in the LP domain is reduced, selective normalization improves the correlation coefficient between the original and recovered watermarks.

3.1. Neyman-Pearson based watermark detector

A Neyman-Pearson solution to watermark detection is applicable to the binary decision problem of determining the presence or absence of a particular watermark vector in the received signal distorted by additive noise. Preliminary experiments are used to set the hypotheses,

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



(a) (b)

Fig. 1. Noise distribution in the LP domain.

where $\{r_i\}_{i=1}^{L}$ is the observation vector. The null hypothesis is that no watermark is present and only noise is transmitted $\{v_i\}_{i=1}^{L}$, while under H_1 both the watermark $\{\omega_i\}_{i=1}^{L}$ and noise samples $\{v_i\}_{i=1}^{L}$ are present in additive combination. Due to selective normalization of watermark coefficients, the noise in LP domain, v_i is approximately distributed as $\mathcal{N}(0, \sigma^2)$, when noise $\{\zeta_i\}_{i=1}^{N}$ is added to the stegosignal in the time domain such that the SNR is $S_1 = 10 \log_{10} \frac{\sum_{n=1}^{N} \tilde{y}_n^2}{\sum_{n=1}^{N} \zeta_n^2}$. For this watermark detection problem, the expressions for false-alarm, detection and missed detection rates are well-known and are given by [9],

$$P_F = 0.5 \left[\mathbf{erfc} \left(\frac{\ln \eta + \sum_{i=1}^{L} \frac{\omega_i^2}{2\sigma^2}}{\sqrt{2}\bar{\sigma}} \right) \right] \tag{6}$$

$$P_D = 0.5 \left[\mathbf{erfc} \left(\frac{\ln \eta + \sum_{i=1}^{L} \frac{\omega_i^2}{2\sigma^2} - \bar{\mu_1}}{\sqrt{2}\bar{\sigma}} \right) \right] \tag{7}$$

$$P_M = 1 - P_D \tag{8}$$

Here, $\bar{\mu_1} = \sum_{i=1}^{L} \frac{\omega_i^2}{\sigma^2}$, $\bar{\sigma} = \sqrt{\sum_{i=1}^{L} \frac{\omega_i^2}{\sigma^2}}$ and η is the detection threshold. Let $\eta'' = \ln \eta + \sum_{i=1}^{L} \frac{\omega_i^2}{2\sigma^2}$ then, the decision rule is

$$\sum_{i=1}^{L} \frac{r_i \omega_i}{\sigma^2} \begin{array}{c} H_1 \\ \geq \\ < \\ H_0 \end{array} \eta''. \tag{9}$$

In a practical implementation, the threshold τ'', corresponding to an SNR of S_1, can be adjusted further if the actual SNR in the time domain is determined. As an example, if the SNR was found to be $S_2 = 10 \log_{10} \frac{\sum_{n=1}^{N} \tilde{y}_n^2}{\sum_{n=1}^{N} \zeta_n^2}$ (assuming zero-mean noise), the threshold τ'' is altered by multiplying σ^2 with the adjustment factor $1/\beta$, where $\beta = 10^{(\frac{S_1 - S_2}{10})}$. Equation (9) is based on the assumption that the detector receives watermarked and unwatermarked signals with equal probability.

The SNR in the parametric domain is given by, $d^2 = (\frac{\bar{\mu_1}}{\bar{\sigma}})^2$ [9]. In the prsent case $d = \sqrt{\bar{\mu_1}}$. Embedded marks of greater energy will result in improved robustness, while noise of higher variance in the parametric domain will hinder watermark detection. The stegosignal was subjected to white and colored noise, resulting in different SNRs in the time and parametric domain. In each case, experiments were repeated 1000 times in order to estimate the mean and variance of the Gaussian noise affecting each watermark coefficient. The receiver operating characteristics (ROC) were determined using equations (6) and (7). It can be observed from Table 4 that very low false-positive rates can be obtained for parametric watermarking with selective normalization. When 10 dB white

Table 4. ESTIMATES OF SNR, d^2, P_D AND P_F

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

Table 5. MINIMUM-ERROR-RATE DETECTION USING DISCRIMINANT FUNCTIONS

Noise	SNR (dB)	$\mathbf{P_D}$	$\mathbf{P_F}$
White	3dB	0.9289	0.2487
White	7dB	0.9939	0.0797
White	10dB	0.9997	0.0198
White	15dB	0.9998	$10\times^{-5}$

noise is added to the stegosignal, for a threshold $\tau'' = 4.3960$, a $P_D = 0.99999$ and a false-alarm rate $P_F = 1.37 \times 10^{-6}$ are obtained. Experiments for time domain SNRs of 1 dB and 3 dB resulted in P_F of 0.14 and 0.0033 respectively, an improvement over the results in Table 4.

3.2. Discriminant functions for watermark detection

Discriminant functions [10] are used to assign the recovered vector to one of the many possible classes. In the present application, there are nine classes which include the eight orthogonal watermarks and the case of no watermark being present. Discriminant functions are derived based on minimum-error-rate classification [10] for all the nine classes. The recovered vector is assigned to the class that results in the highest discriminant function.

Detection is done on a per frame basis. The parameter noise affecting a watermark coefficient when colored noise is added in the time domain, can be approximated by $\mathcal{N}(0, \sigma^2)$. The noise distribution was the same even when no watermark was embeded. Since vectors of dimension eight were considered, the noise affecting all the eight coefficients of the recovered vector for a particular class k can be expressed by the general multivariate normal density,

$$p_k(\mathbf{r}) = \frac{1}{(2\pi)^(4)|\mathbf{\Sigma}|^{\frac{1}{2}}} \exp[-\frac{1}{2}(\mathbf{r} - \mu_k)'\mathbf{\Sigma}^{-1}(\mathbf{r} - \mu_k)] \quad (10)$$

where $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, $\mathbf{I}$ being the identity matrix, and μ_k is the mean of the multivariate normal density for class k. For this application, k takes values from 1 to 9, with 9 corresponding to the 'no noise' case. Hence, $\mu_k = \omega^k$ for $k = 1, \cdots, 8$ and $\mu_{k=9} = \mathbf{0}$, with $\mathbf{0}$ being the zero vector. The discriminant function for class k is given by,

$$\mathbf{g}_k(\mathbf{r}) = -\frac{1}{2\sigma^2}[\mathbf{r}'\mathbf{r} - 2\mu_\mathbf{k}'\mathbf{r} + \mu_\mathbf{k}'\mu_\mathbf{k}] \quad (11)$$

The recovered vector $\mathbf{r}$ is assigned to the class the j satisfying the following equation.

$$\mathbf{g}_j(\mathbf{r}) > \mathbf{g}_k(\mathbf{r}) \quad \text{for all} j \neq k. \quad (12)$$

Thus using the above expression, a watermark present in the recovered vector is detected. Experiments were performed to determine the false-alarm and detection rates and the results are tabulated in Table 5. For time domain SNRs below 15 dB, the resulting stegosignals are highly noisy and are of no commercial significance. Due to low false-alarm rates ($< 10^{-6}$), it is not feasible to experimentally determine P_F for SNRs above 15 dB. In most watermarking applications, it is necessary to have very low false-alarm rates. Instead of using minimum-error-rate criterion where both missed detection and false-alarm are equally costly, the false-alarm error can be weighted more heavily and the discriminant functions derived accordingly.

4. CONCLUSIONS

Parameter noise is analyzed when the stegosignal is distorted by additive noise. Selective normalization of watermark coefficients is proposed. By using Neyman-Pearson detector and minimum-error-rate classification, the false-alarm and missed detection rates are determined.

5. REFERENCES

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

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

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

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

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

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

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

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

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

[10] R. O. DUDA, P. E. HART and D. G. STORK, *Pattern Classification* (2d ed.), Wiley Interscience, 2001.