SPEECH WATERMARKING BY PARAMETRIC EMBEDDING WITH AN ℓ_∞ FIDELITY CRITERION

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

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

ABSTRACT

Parameter-embedded watermarking of speech signals is effected through slight perturbations of parametric models of some deeply-integrated dynamics of the signal. One of the objectives of the present research is to develop, within the parameter-embedding framework, quantifiable measures of fidelity of the stegosignal and of robustness of the watermark to attack. This paper advances previous developments on parameter-embedded watermarking by introducing a specific technique for watermark selection subject to a fidelity constraint. New results in set-theoretic filtering are used to obtain sets of allowable parameter perturbations (i.e., watermarks) subject to an ℓ_∞ constraint on the error between the watermarked and original material. With respect to previous trial-and-error perturbation methods, the set-based parameter perturbation is not only quantified and systematic, it is found to be more robust, and to have a higher threshold of perceptibility with perturbation energy. After a brief review of the general parameter-embedding strategy, the new algorithm for set-theoretic watermark selection is presented. Experiments with real speech data are used to assess robustness and other performance properties. This work is being undertaken in support of the development of the National Gallery of the Spoken Word, a project of the Digital Libraries II Initiative.

1. MOTIVATING PROBLEM

The results reported in this paper represent significant enhancements to a speech watermarking technique first presented at the 2002 International Conference on Spoken Language Processing (ICSLP02) [1]. The motivating application for this work is the creation of the *National Gallery of the Spoken Word* (NGSW), a Digital Libraries Initiative II project whose goal is the development of a carefully organized on-line repository of spoken word collections, based largely upon the renowned Vincent Voice Library at Michigan State University (MSU). Speech engineering aspects of the NGSW are being researched by the authors in collaboration with the Center for Spoken Language Research (CSLR) at the University of Colorado. A brief introduction to the NGSW project by CSLR researchers is found in [2], and further information is available at www.ngsw.org.

2. PARAMETER-EMBEDDED WATERMARKS

The general method presented at ICSLP 2002 is referred to as *parameter-embedded watermarking*. The method involves *informed embedding*, meaning simply that the original, unmarked, material – the *coversignal* – (or equivalent information) is required to recover or detect the watermark from the marked material – the *stegosignal*. Generally speaking, parameter-embedded watermarking is effected through the introduction of slight perturbations into parametric models of some deeply-integrated temporal aspect of the signal. This idea can be formalized in broadly general terms, but in the present research, we adopt the following formulation: Let $\{y_t\}$ denote the coversignal (in particular, a frame of the coversignal that is to be watermarked), and let $\{\bar{y}_t\}$ be the stegosignal. Each is a real scalar sequence over discrete-time, t. It is assumed that these signals are generated according to operations of the form

$$y_t = \Phi_\pi\left(\xi_t, \mathbf{x}_t, t\right) \quad \text{and} \quad \bar{y}_t = \phi_{\bar{\pi}}\left(\bar{\xi}_t, \bar{\mathbf{x}}_t, t\right) \tag{1}$$

in which $\{\xi_t\}$, $\{\bar{\xi}_t\}$, $\{\mathbf{x}_t\}$, and $\{\bar{\mathbf{x}}_t\}$ are measurable vector-valued random sequences. The operator Φ is parameterized by a set π, the alteration of which (to parameter set $\bar{\pi}$) is responsible for changing the operator Φ into ϕ and sequences $\{\xi_t\}$ and $\{\mathbf{x}_t\}$ into their "overbarred" counterparts. In the earlier work, the coversignal was assumed to be generated by an autoregressive (AR), or linear prediction (LP), model,

$$y_t = \sum_{i=1}^{M} a_i y_{t-i} + \xi_t \tag{2}$$

but this "true" model was actually estimated from the speech frame using standard LP analysis (e.g. [3]). The sequence $\{\xi_t\}$, therefore, is the (known) prediction residual associated with the estimated model. The stegosignal was constructed using the FIR filter

$$\bar{y}_t = \sum_{i=1}^{M} \bar{a}_i y_{t-i} + \xi_t \tag{3}$$

where $\{\bar{a}_i\}$ represents a slightly perturbed version of the "true" set $\{a_i\}$. The perturbation is effected via the autocorrelation values in the normal equations. The earlier technique was found to be remarkably robust to a wide variety of attacks, including cropping, additive noise, MP3 compression, and others. However, the "watermarks" in the form of model parameter perturbations were necessarily constructed by trial-and-error manipulation of the first few autocorrelation values. Such a procedure offers little possibility of quantifying relationships between watermarking strategy and performance measures. The work reported in this paper represents a step toward such quantification.

This work was supported by the National Science Foundation of the U.S. under Cooperative Agreement IIS-9817485. Opinions expressed are those of the authors and do not necessarily reflect the views of the NSF.

3. ℓ_∞ FIDELITY CRITERION

Problem Statement. The design of a watermarking strategy involves the balancing of two principal criteria. First, watermarks must be imperceptible to the listener (or viewer). Second, watermarks must be robust. That is, they must be able to survive *attacks* – those deliberately designed to destroy or remove them, as well as distortions inadvertently imposed upon the watermarks by specific technical processes (e.g., compression) or by systemic processes like channel noise or computational roundoff errors. The fidelity and robustness criteria are generally competing in the sense that greater robustness requires more watermark energy, more manipulation of the coversignal, etc., which, in turn, ultimately leads to noticeable distortion of the original content. Related measures of a watermark's efficacy include *data payload*, which refers to the number of watermark bits within a unit of time or work, and its *security*, the inherent ability of the system to prevent unauthorized removal, embedding or detection [4].

As a step toward quantifying the relationships between the two competing performance measures and the watermarking strategy, the following general problem is addressed in this research:

CONSTRAINED WATERMARKING PROBLEM. *For coversignal frame $\{y_t\}_{t=1}^T$ generated according to model (2) with parameters π, find the set of watermarks (i.e., the set of parameter sets $\{\bar{\pi}\}$) such that, for stegosignal frame $\{y_t\}_{t=1}^T$ generated according to (3), the following fidelity criterion is met,*

$$\|\mathbf{y} - \bar{\mathbf{y}}\|_\infty < \gamma \tag{4}$$

in which $\mathbf{y}$ and $\bar{\mathbf{y}}$ are T-vectors with ith elements y_i and $\bar{y}_i$, respectively.

FORMULATION USED IN PRESENT WORK. *The specific instance of this problem represented by the generation equations (2) and (3) is addressed here, wherein $\pi = \{a_i\}_{i=1}^M$, $\bar{\pi} = \{\bar{a}_i\}_{i=1}^M$, $\mathbf{x}_t = \bar{\mathbf{x}}_t = \begin{bmatrix} y_{t-1} & \cdots & y_{t-M} \end{bmatrix}^T$, and $\bar{\xi}_t = \xi_t$. Φ_π and $\phi_{\bar{\pi}}$ represent the AR and FIR filter operations (2) and (3), respectively.*

Set-Membership Filtering. The solution to the problem posed above follows from a recent result from adaptive filtering research known as *set-membership filtering* (SMF). The SMF concept was first published by Gollamudi *et al.* [5], and was more recently proposed as an innovative solution to the design of channel equalizers for digital communication by Nagaraj *et al.* [6]. SMF can be viewed as a reformulation of the broadly-researched class of algorithms concerned with *set-membership identification* (e.g. [7, 8]). The SMF problem is stated as follows:

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

$$\mathcal{P}_t = \left\{\theta \,\middle|\, |z_\tau - \theta^T \mathbf{x}_\tau| < \gamma_\tau \text{ for } \tau \in [1, t]\right\}. \tag{5}$$

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

$$\mathcal{P}_t = \left\{\theta \,\middle|\, \|\mathbf{z} - \bar{\mathbf{z}}\|_\infty < \gamma\right\}. \tag{6}$$

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

The SMF problem is solved using a series of recursions which ultimately return an hyperellipsoidal membership set, say $\mathcal{E}_t \supset \mathcal{P}_t$, and the ellipsoid's center, say θ_t. The recursions execute an optimization strategy designed to tightly bound $\mathcal{P}_t$ by $\mathcal{E}_t$ in some sense. Accordingly, the broad class of algorithms employed in the SMF problem are often called the *optimal bounding ellipsoid* (*OBE*) algorithms.[1] Results on the theory and application of OBE algorithms abound, and the reader is referred, for example, to the following tutorial papers as an entry points into the literature: [7, 8, 9, 10, 11]. The OBE algorithm used in the experiments to follow is called the *set-membership—weighted recursive least squares* (SM-WRLS) algorithm, but the choice is somewhat arbitrary for the present application.

Adapting SMF to the Constrained Watermarking Problem. The specific watermarking problem posed for the present work is readily solved as an SMF problem as follows. Let

$$\bar{\mathbf{a}} = \begin{bmatrix} \bar{a}_1 & \cdots & \bar{a}_M \end{bmatrix}^T$$

denote a general vector of perturbed filter parameters as in (3). For time frame $t = 1, \ldots, T$, then, the objective is to find the set

$$\mathcal{P}_T = \left\{\bar{\mathbf{a}} \,\middle|\, \|\mathbf{y} - \bar{\mathbf{y}}(\bar{\mathbf{a}})\|_\infty < \gamma\right\} \tag{7}$$

in which $\mathbf{y}$ and $\bar{\mathbf{y}}$ are the usual T-vectors of cover- and stegosignal samples. Of course, it is sufficient to assure that $|y_t - \bar{y}_t| < \gamma$ for each t in the frame. Let us now subtract y_t from each side of (3), multiply through by (-1), then rearrange to obtain

$$y_t - \bar{y}_t = (y_t - \xi_t) - \sum_{i=1}^M \bar{a}_i y_{t-i} = (y_t - \xi_t) - \bar{\mathbf{a}}^T \mathbf{x}_t. \tag{8}$$

Upon defining the sequence $\{z_t = y_t - \xi_t\}_{t=1}^T$ (recall that $\{\xi_t\}$ is known) and specifying the constraint $|y_t - \bar{y}_t| < \gamma$ for each $t \in [1, T]$, the search for the constrained watermark parameters is reduced to a SMF problem as in (5).

4. EXPERIMENT & DISCUSSION

4.1. Methods & Results

An experiment is used to demonstrate the performance benefits of the new approach. The coversignal is the utterance of the word "dark" from the sentence "She had your dark suit in greasy wash water all year." spoken by a female in the TIMIT database [12]. The coversignal frame consists of 3,001 samples taken at a 16 kHz rate (0.1876 seconds). A single watermark was in the frame using both the new algorithm with fidelity constraints and the predecessor parameter-embedding method presented at ICSLP 2002 [1]. We refer to these as the "SMF-based parameter-embedded watermarking technique" (fidelity-constrained parameter embedding) and the "parameter-embedded watermarking or ICSLP02" (parameter embedding via finessed perturbations of the model parameters) approaches. An $M = 8$ order model is employed in all filtering operations, implying eight parameters with which to encode the watermark.

[1] The algorithm cited above due to Gollamudi, Nagaraj *et al.* [5, 6] is sometimes referred to as "*quasi*-OBE" (QOBE) [a.k.a "BEACON" or "SM-NLMS"] because its operation has useful interpretations which do not involve the ellipsoid.

Parameter-embedded watermarking. The Levinson-Durbin recursion (e.g., [3, Ch.5]) was used to obtain LP parameters [see model (2)] for the coversignal, $\{a_i\}$, and the accompanying prediction residual $\{\xi_t\}$. The stegosignal model parameters [see model (3)], $\{\bar{a}_i\}$, were obtained indirectly by adding the watermark sequence, $\{0, 0, 4, 1, 1, 3, 1, 1\} \times 0.01$ to the corresponding autocorrelation sequence points. The 0.01 factor was necessary to scale down the watermark to ensure imperceptibility. The stegosignal was reconstructed according to model (3). A straightforward procedure described in [1, Table 2] was used to recover the watermark. The ℓ_∞ norm of the cover / stegosignal error over the analysis frame was $\|\mathbf{y} - \bar{\mathbf{y}}\|_\infty = 0.917$. The maximum absolute difference between the cover- and stegosignals was found to be 0.917. The cover- and the stegosignals obtained using the ICSLP02 algorithm are shown in Figs. 1(a) and (b).

SMF-based parameter-embedded watermarking. For this approach, the error bound, γ [see (7)] was taken to be a constant 0.917, corresponding to the worst case error for the ICSLP02 approach. From the SM-WRLS OBE algorithm [recall discussion of SMF problem in Sec. 3], a set of allowable "perturbed" solutions satisfying the fidelity constraint were obtained in light of the observations [recall (8) and foregoing discussion]. The central estimate of the hyperellipsoid is taken to be the perturbed parameter solution $\{\bar{a}_i\}_{i=1}^{M}$ (inherently containing the watermark) for use in (3). The difference between the autocorrelation values corresponding to the true LP coefficients and the autocorrelation values corresponding to the perturbed-parameter filter was computed (nominal watermark for comparison with ICSLP02 watermark). The watermark was taken to be the sequence of numbers obtained on multiplying this difference by a factor of 10 and rounding off, yielding, in the present experiment $\{6, 5, 3, 2, 0, 1, 2, 2\}$.

It is observed that a more energetic watermark sequence is obtained in this case. Experiments have demonstrated that the strength of the watermark plays a key role in the robustness to common types of attacks. An improvement in robustness, in terms of increased watermark energy can be obtained at the expense of stegosignal fidelity. Further $\|\mathbf{y} - \bar{\mathbf{y}}\|_\infty = 0.682$, is significantly less than the prescribed constraint $\gamma = 0.917$. This reduction in the worst-case error has a significant, positive, impact on the fidelity of the reconstructed stegosignal [*c.f.* Figs. 1(b) and (c)].

4.2. Discussion

Fidelity. The principal objective of the present work was to demonstrate a method by which the watermarking procedure could be quantitatively related to the first of two competing objectives, fidelity and robustness. In limited, informal listening tests by the authors, the stegosignal resulting from the SMF-based watermarking approach was consistently more like the coversignal perceptually than that resulting from the ICSLP02 algorithm. This was a general observation by the authors and no formal subjective tests were performed to assess stegosignal fidelity. Indeed, the experimental fidelity "measure," $\|\mathbf{y} - \bar{\mathbf{y}}\|_\infty = 0.682$, was significantly better than the criterion set for this measure, $\gamma = 0.917$. The ability to systematically compute a set of perturbed filters that are consistent with the data [from which to chose the "stegosignal generator" (3)] is apparently very significant in preserving fidelity to the original coversignal. This is not surprising since the perturbed parameters are selected from a set which contains the "true" parameters, and whose elements are analytically meaningfully tied through observations to the structure of the signal. This is juxtaposed with the

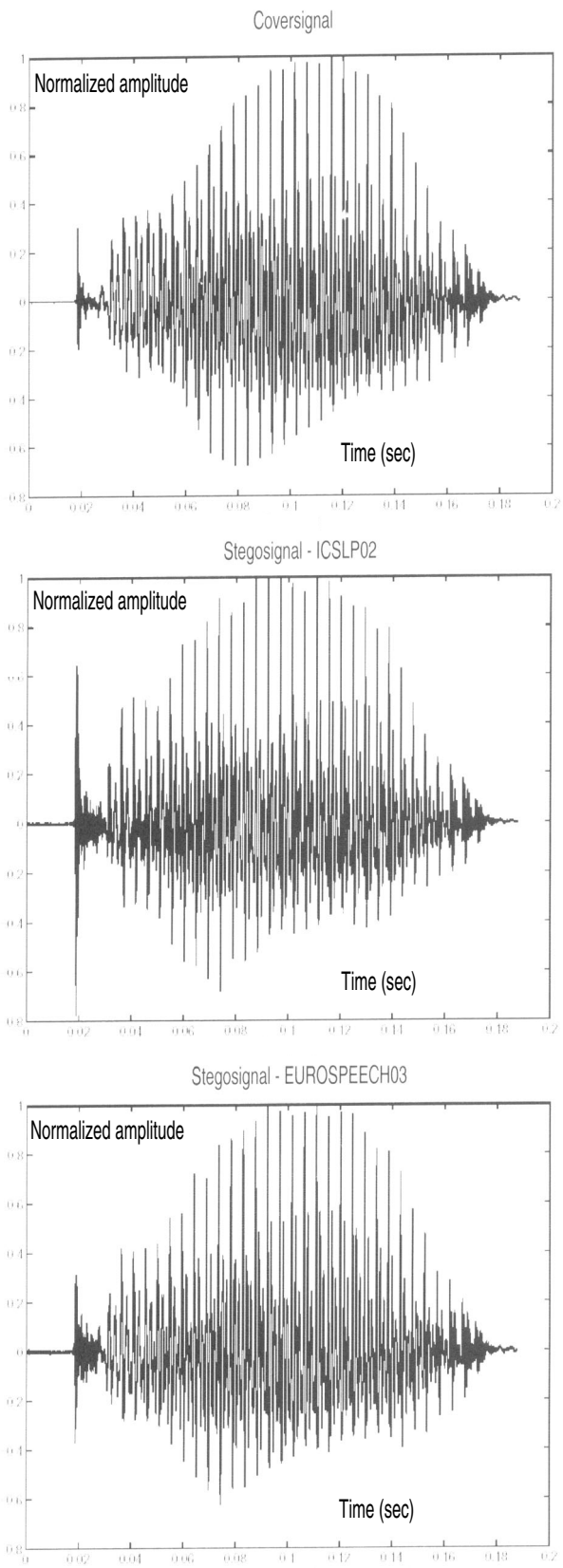

Fig. 1. (a) Coversignal frame. (b) Stegosignal frame resulting from the ICSLP02 watermarking method. (c) Stegosignal frame resulting from the SMF-based parameter-embedded watermarking method with ℓ_∞ fidelity constraint.

more *ad hoc* approach taken in ICSLP02 wherein the precise effects on the stegosignal filter caused by "manual" perturbations of the autocorrelation sequence are unknown.

The value of $\gamma = 0.917$ was selected for the SMF-based parameter-embedded watermarking experiment above to provide parity in comparing the two techniques. In practice, γ must be selected according to the fidelity requirements of the intended application. Indeed, it is the ability to quantitatively prescribe fidelity requirements that is the essential contribution of this work.

Robustness. The parameter-embedded watermarking algorithm exhibits significant ability to withstand most of the common attacks to which watermarks are subjected [1]. One reason for the greater robustness is that the watermark signal is concentrated during the embedding and recovery processes, while it is spread across the entire work otherwise [4]. The SMF-based version of the algorithm results in improved performance by embedding more energetic watermarks. In preliminary experiments, the improvement in fidelity afforded by SMF-based watermarking is accompanied by improved robustness performance. As an example, ICSLP02 was robust against jitter attack when one in every 12 samples was duplicated or when one out of every 30 samples was set to zero [1]. In a similar implementation of the jitter attack for the SMF-based technique, the embedded watermark was recovered when one in eight samples was duplicated, or when one of every 38 samples was randomly set to zero. The reasons for this improvement involves the fact that the SMF-based algorithm is able to preserve fidelity with a much more energetic watermark than that which can be used in ICSLP02. Further, the technique by which parameters are perturbed yields watermarking filters which are heuristically "closer" (in terms of consonance with the observed data) than those which necessarily arise from the ICSLP02 method. This might mean that the SMF-based filters are preserving more perceptually-important spectral information (consistent with higher fidelity) which is not as readily attacked without perceptual damage.

5. CONCLUSIONS

This paper has described a technique for integrating a quantified specification of fidelity into a watermarking strategy based on perturbed parameter watermarking. In particular, recent results in set-theoretic filtering are used to obtain sets of allowable parameter perturbations (i.e., watermarks) subject to an ℓ_∞ constraint on the error between the watermarked and original material. With respect to previous trial-and-error perturbation methods, the set-based parameter perturbation is not only quantified and systematic, it is found to be more robust than the former, and to have a higher threshold of perceptibility with perturbation energy. Much work remains to explain the improved performance of the new method, and to seek quantification of the robustness requirements as well. Further results on the quantitative relationship between watermark robustness and stegosignal fidelity will appear in [13]. One of the most novel aspects of the present method is the use of a solution set of watermarks. Since location of the perturbed parameter vector within the solution set is intricately tied to spectral properties of the signals, exploitation of set properties remains an interesting issue for further research.

6. REFERENCES

[1] A. GURIJALA, J.R. DELLER, JR. and M.S. SEADLE "Speech watewatermarking through parametric modelling," *Proc. Int'l. Conf. Spoken Language Processing*, Denver, Sep. 2002, on CD-ROM.

[2] J.H.L. HANSEN, B. ZHOU, ET AL., "Audio stream phrase recognition for a National Gallery of the Spoken Word: One small step," *Proc. Int'l. Conf. Spoken Language Processing*, Beijing, 1089-1092, Oct. 2000.

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

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

[5] S. GOLLAMUDI, S. NAGARAJ, S. KAPOOR and Y.F. HUANG, "SMART: A toolbox for set-membership filtering," *Proc. 1997 European Conf. Circuit Theory and Design*, Budapest, 1997.

[6] S. NAGARAJ, S. GOLLAMUDI, S. KAPOOR and Y.F. HUANG, "BEACON: An adaptive set-membership filtering technique with sparse updates," *IEEE Trans. on Signal Processing*, vol. 47, pp. 2928–2941, 1999.

[7] J.R. DELLER, JR. and Y.F. HUANG, "Set-membership identification and filtering in signal processing," *Circuits, Systems, and Signal Processing* [Special issue on signal processing and its applications], Feb. 2002.

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

[9] M. MILANESE, J.P. NORTON, H. PIET-LAHANIER and E. WALTER (eds.), *Bounding Approaches to System Identification*, London: Plenum, 1996.

[10] J.P. NORTON (guest ed.), *Int. J. Automatic Control and Signal Process.*, Two-part special issue on bounded error methods in control and signal processing, vol. 8 (issues 1 & 2), 1994.

[11] E. WALTER (ed.), *Mathematics and Computers in Simulation*, Special issue on parameter identification with error bound, vol. 32, 1990.

[12] NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY (NIST) "Getting started with the DARPA TIMIT CD-ROM," Gaithersburg MD, 1988.

[13] A.R. GURIJALA and J.R. DELLER, JR. "Speech watermarking with objective fidelity and robustness criteria," *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove CA, Nov. 2003 (in press).