EXPONENTIAL WEIGHTING METHOD FOR SAMPLE-BY-SAMPLE UPDATE OF WARPED AR-MODEL

Kari Roth

University of Turku / Department of Physics
Laboratory of Optics and Spectroscopy
FIN-20014 Turun Yliopisto, Finland
kari.roth@utu.fi

Ismo Kauppinen

Noveltech Solutions ltd
Tykistökatu 4
FIN-20520 Turku, Finland
ismo@noveltechsolutions.com

ABSTRACT

Auto-regressive (AR) modeling is a powerful tool having many applications in audio signal processing. The modeling procedure can be focused to low or high frequency range using frequency warping. Conventionally the AR-modeling procedure is accomplished with frame-by-frame processing which introduces latency. As with any frame-by-frame algorithm full frame has to be available for the algorithm before any output can be produced. This latency makes AR-modeling more or less unusable in real-time sound effects especially when long frame lengths are required. In this paper we introduce an exponential weighting (EW) method for sample-by-sample update of the warped AR-model. This method reduces the latency down to the order of the AR-model.

1. INTRODUCTION

An auto-regressive model [1] assumes that a signal sample y_n can be predicted as a linear combination of previous samples and is defined by

$$y_n = -\sum_{m=1}^{p} a_m y_{n-m} + e_n, \tag{1}$$

where p is the order of the model, a_m are the model coefficients, and e_n is the prediction error (i.e. residual). When signal samples are known the prediction error is given by

$$e_n = y_n + \sum_{m=1}^{p} a_m y_{n-m}. \tag{2}$$

The model coeffients are chosen to minimize the prediction error usually in a least squares sense. There exist several algorithms for calculating the model coefficients (see e.g. [2]). In this paper Burg's method [3] is used and a time-domain exponential weighting (EW) method for sample-by-sample update of the model coefficients is introduced.

Frequency warping [4] can be used to emphasize the modeling procedure to the low or high frequency range of the signal. This is highly useful in audio signal processing since audio signals usually have most of their energy concentrated to frequencies below $10\,\text{kHz}$. Frequency warping is achieved via bilinear mapping in the z-plane

$$\tilde{z}^{-1} = D(z) = \frac{z^{-1} - \lambda}{1 - \lambda z^{-1}}, \tag{3}$$

where λ is the warping parameter. This implies that frequency warping can be incorporated into existing methods simply by replacing unit delays z^{-1} by allpass filters $D(z)$.

2. WARPED BURG'S METHOD

Equation (2) represents a finite impulse response (FIR) filter and therefore can also be realized by a lattice structure depicted in Figure 1. The reflection coefficients are chosen to minimize the forward and backward prediction errors $f_n^{(m)}$ and $b_n^{(m)}$ at each stage independently. Solution to this problem is given as

$$k_m = -\frac{2\langle f_n^{(m-1)} b_{n-1}^{(m-1)} \rangle}{\langle f_n^{(m-1)} f_n^{(m-1)} \rangle + \langle b_{n-1}^{(m-1)} b_{n-1}^{(m-1)} \rangle}, \tag{4}$$

where $\langle\ \rangle$ denotes expectation value. This reflection coefficient is then used to obtain prediction errors of stage m

$$\begin{aligned} f_n^{(m)} &= f_n^{(m-1)} + k_m b_{n-1}^{(m-1)} \\ b_n^{(m)} &= b_{n-1}^{(m-1)} + k_m f_n^{(m-1)}. \end{aligned} \tag{5}$$

Initially $f_n^{(0)} = b_n^{(0)} = y_n$. If the actual model coefficients are needed they can be obtained via Levinson-Durbin recursion [5, 6]. When frequency warping is used in Burg's method the unit delays of the lattice filter are replaced by warped allpass filters $D(z)$. The resulting structure is shown in Figure 2. Frequency warping causes only minor modifications to the modeling equations. The warped reflection coefficient is given by

$$\tilde{k}_m = -\frac{2\langle f_n^{(m-1)} \tilde{b}_n^{(m-1)} \rangle}{\langle f_n^{(m-1)} f_n^{(m-1)} \rangle + \langle \tilde{b}_n^{(m-1)} \tilde{b}_n^{(m-1)} \rangle}, \tag{6}$$

where $m = 1, 2, \ldots, p$ and

$$\tilde{b}_n^{(m)} = b_{n-1}^{(m-1)} - \lambda \left[b_n^{(m-1)} - \tilde{b}_{n-1}^{(m)} \right], \tag{7}$$

which is the warped backward prediction error. The reflection equations are respectively transformed to

$$\begin{aligned} f_n^{(m)} &= f_n^{(m-1)} + \tilde{k}_m \tilde{b}_n^{(m)} \\ b_n^{(m)} &= \tilde{b}_n^{(m)} + \tilde{k}_m f_n^{(m-1)} \end{aligned} \tag{8}$$

For a more detailed description see [7].

3. SAMPLE-BY-SAMPLE UPDATE

When using frame-by-frame AR-modeling the latency of the modeling phase is at least equal to the frame size used. This causes a disturbing delay if AR-modeling is used e.g. in a real-time sound

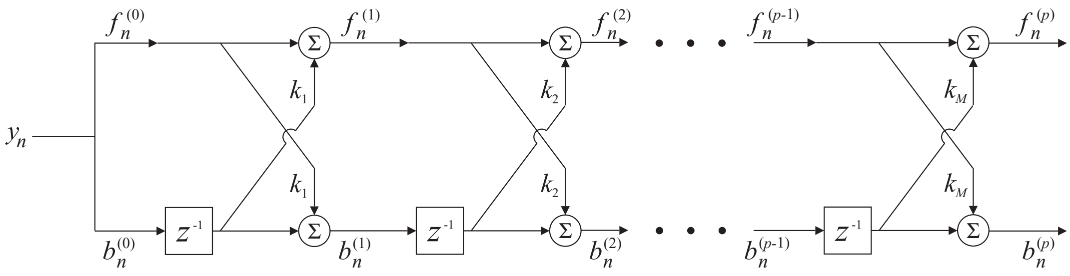

Figure 1: *The lattice filter structure used in conventional Burg's method.*

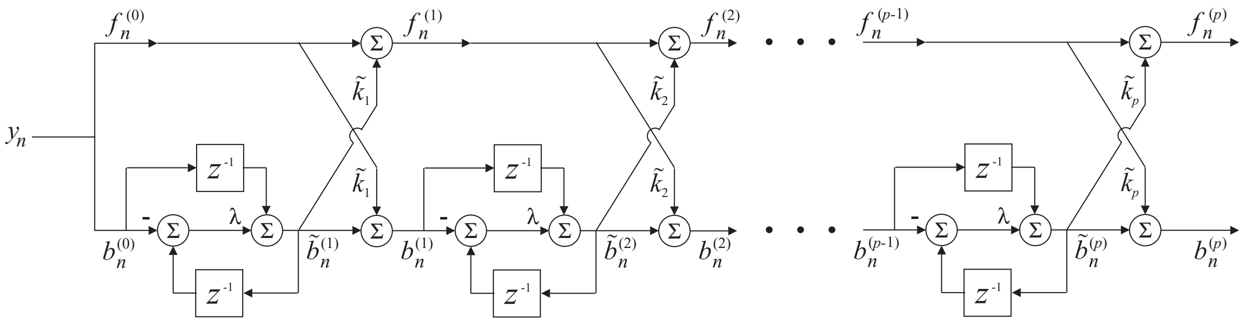

Figure 2: *The lattice filter structure used in warped Burg's method.*

effect. By using sample-by-sample update this latency can be reduced to be equal to the order of the model.

Previously, a method has been developed for obtaining sample-by-sample update of the AR-model coefficients in a non-warped case, namely the gradient adaptive lattice method (GAL) [8]. This method can also be further developed to the frequency warped case. However, GAL does not ensure the stability of the model, i.e. k_m are not guaranteed to have absolute value less than unity although usually this is the case.

3.1. Gradient adaptive lattice method

GAL method uses the gradient of the prediction error energy

$$\hat{J}_n^{(m)} = \left[f_n^{(m)}\right]^2 + \left[b_n^{(m)}\right]^2. \tag{9}$$

The updated reflection coefficient is

$$k_m(n) = k_m(n-1) + \Delta k_m(n), \tag{10}$$

where $\Delta k_m(n)$ can be estimated by using steepest descent approach which gives

$$\Delta k_m(n) = -\mu_m \frac{\partial \hat{J}_n^{(m)}}{\partial k_m}, \tag{11}$$

where μ_m is a step size parameter and

$$\frac{\partial \hat{J}_n^{(m)}}{\partial k_m} = 2\left[f_n^{(m)} b_{(n-1)}^{(m-1)} + b_n^{(m)} f_n^{(m-1)}\right]. \tag{12}$$

The step size parameter μ_m can be adapted to the prediction error energy estimate to have faster adaptation for fast changes in signal. This is done by choosing

$$\mu_m = \frac{1-\beta}{2\hat{D}_n^{(m)}} \tag{13}$$

where β is a smoothing parameter and $\hat{D}_n^{(m)}$ is the estimate of the prediction error energy after lattice stage m having value

$$\hat{D}_n^{(m)} = \beta \hat{D}_{n-1}^{(m)} + (1-\beta)\left\{\left[f_n^{(m-1)}\right]^2 + \left[b_{n-1}^{(m-1)}\right]^2\right\}. \tag{14}$$

3.2. Exponential weighting method

The proposed EW method for sample-by-sample update for the model parameters is to use time-domain exponential weighting to calculate the expectation values in equation (6). This can be achieved by

$$\begin{aligned}
\langle f_n^{(m)} f_n^{(m)} \rangle &\approx F_n^{(m)} = \alpha F_{n-1}^{(m)} + (1-\alpha) f_n^{(m)} f_n^{(m)} \\
\langle \tilde{b}_n^{(m)} \tilde{b}_n^{(m)} \rangle &\approx B_n^{(m)} = \alpha B_{n-1}^{(m)} + (1-\alpha) \tilde{b}_n^{(m)} \tilde{b}_n^{(m)} \\
\langle f_n^{(m)} \tilde{b}_n^{(m)} \rangle &\approx X_n^{(m)} = \alpha X_{n-1}^{(m)} + (1-\alpha) f_n^{(m)} \tilde{b}_n^{(m)},
\end{aligned} \tag{15}$$

where α is a smoothing parameter. The higher the value of α is the more weight is given to the past values and the longer is the time required for the model to adapt to changes in the source. The time constant of the adaptation is

$$\tau = \frac{1-\alpha}{\alpha} \Delta t, \tag{16}$$

Table 1: *Comparison of the operations needed per sample in each lattice stage for gradient adaptive lattice method, proposed EW method and EW with warping*

	ADD	MULT	DIV
GAL	8	11	1
EW	7	11	1
EW + Warping	9	12	1

where Δt is the sampling interval. Now the reflection coefficient $\tilde{k}_m$ can be calculated from

$$\tilde{k}_m(n) = -\frac{2X_n^{(m)}}{F_n^{(m)} + B_n^{(m)}} \quad (17)$$

3.3. Computational costs

In Table 1 the computational costs of GAL method, the proposed EW method, and EW method with warping are presented. The computational costs of GAL and EW methods are nearly equal. The computational costs for GAL includes step size adaptation.

No quantitative comparison between frame-by-frame methods and sample-by-sample methods are shown here. Usually block processing methods have to use overlap-and-add which causes each sample to be processed at least twice making the computational costs greater than either of the sample-by-sample methods discussed in this paper.

3.4. Modeling performance

In Figures 3 and 4 the modeling performances of the EW method, GAL method, and the frame-by-frame method are compared. The modeled signal was a guitar tone sampled at $44.1\,\text{kHz}$. Model order p was 6. The figures represent the tracks of reflection coefficients k_2, k_3, and k_4 as a function of time. The smoothing parameter α had value of 0.9999 in both experiments

Figure 3 compares the EW method and the frame-by-frame method. Value $\lambda = 0.2$ was used for warping parameter. With frame-by-frame method frame size 1000 was used with $90\,\%$ overlap. At each position the center of the frame used in frame-by-frame method is aligned with the sample that is fed into the EW algorithm. Figure 4 compares the EW method and the GAL method. In this case no warping was used. From these experiments it is easy to notice that the model obtained with the EW method is consistent with the model obtained using the other methods.

4. APPLICATIONS

AR-modeling has previously been succesfully applied to e.g. partial tracking [9], transient sound analysis [10], click detection [11], and vocoding. With the proposed method frequency warping can be incorporated into these effects without losing real-time nature due to the long latency introduced by block processing.

In AR-model based click detection using frequency warping with negative warping factor enhances the performance [12]. In Figure 5 the proposed method is applied to click detection. The signal sample consist of 50000 samples of orchestral music sampled at 44.1 kHz and corrupted by crackle. The signal is fed through the algorithm and the resulting residual is shown in the lower figure. Model order $p = 6$, warping factor $\lambda = -0.5$, and smoothing parameter value $\alpha = 0.9999$ was used. In the original signal the degradations are well hidden. The modeling process enhances the disturbances and allows a restoration algorithm to focus only on the damaged sections that can be identified e.g. by simple thresholding.

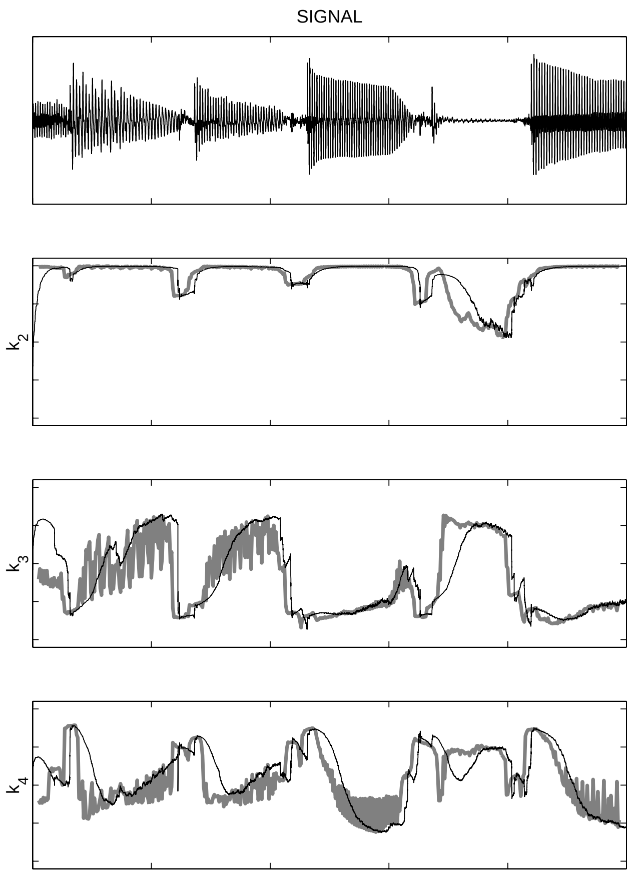

Figure 3: *Tracks of the reflection coefficients using frame-by-frame method (thick gray line) and the exponential weighting method (thin black line).*

5. CONCLUSIONS

In this paper we have adressed the problem of reducing latency in the AR-modeling procedure where frame-by-frame methods are conventionally used. We have introduced an exponential weighting (EW) method for sample-by-sample update of the AR-model including frequency warping. We have shown the calculation efficiency of the method and compared it to the excisting gradient adaptive lattice (GAL) in the non-warped case. The latency is reduced to the order of the model in the proposed method.

6. REFERENCES

[1] S. Haykin, Ed., *Nonlinear Methods of Spectral Analysis*, Springer-Verlag, Berlin, 2nd edition, 1983.

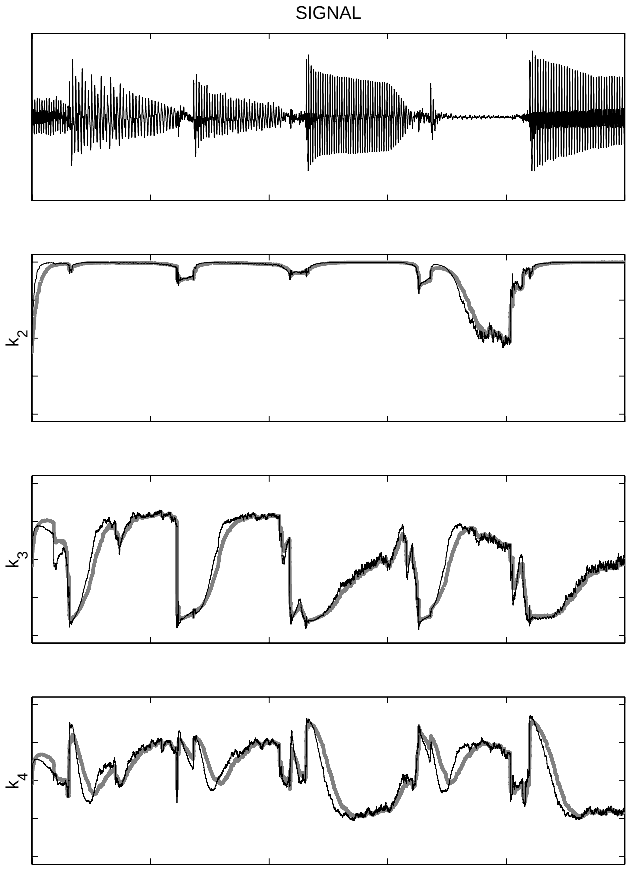

Figure 4: *Tracks of the reflection coefficients using gradient adaptive lattice method (thick gray line) and the exponential weighting method (thin black line).*

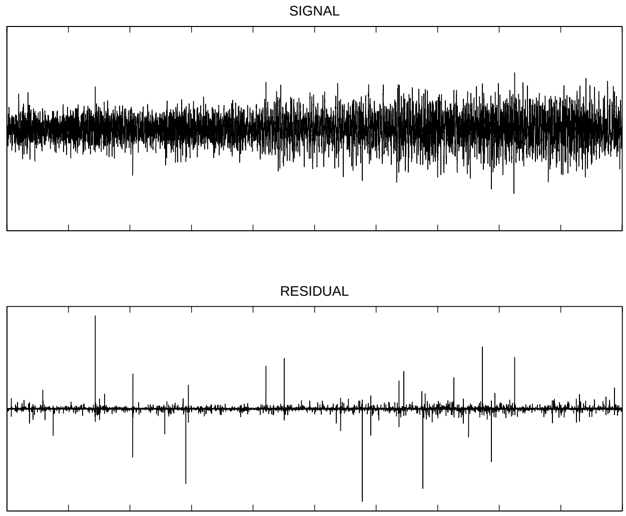

Figure 5: *Signal corrupted by crackle and a residual obtained from it. The employed setup was* $p = 6$, $\lambda = -0.5$, *and* $\alpha = 0.9999$.

[2] J. Makhoul, "Linear Prediction: A Tutorial Review," in *Modern Spectrum Analysis*, D. G. Childers, Ed., pp. 98–118. IEEE Press, New York, 1978.

[3] J. P. Burg, "A New Analysis Technique for Time Series Data," NATO Advanced Study Institute on Signal Processing with Emphasis on Underwater Acoustics, Enschede, The Netherlands, Aug. 1968, Reprinted in *Modern Spectrum Analysis*, D. G. Childers, ed., IEEE Press, New York, 1978.

[4] A. Härmä, M. Karjalainen, V. Välimäki, L. Savioja, U. Laine, and J. Huopaniemi, "Frequency-Warped Signal Processing for Audio Applications," *J. Audio Eng. Soc.*, vol. 48, no. 11, Nov. 2000.

[5] N. Levinson, "The Wiener RMS error criterion in filter design and prediction," *J. Math. Phys.*, vol. 25, pp. 261–278, 1947.

[6] J. Durbin, "The Fitting of Time Series Models," *Rev. Int. Stat. Inst.*, pp. 233–244, 1960.

[7] Kari Roth, Ismo Kauppinen, Paulo Esquef, and Vesa Välimäki, "Frequency Warped Burg's Method for AR-Modeling," in *Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03)*, New Paltz, New York, Oct. 2003.

[8] Florian Keiler, Daniel Arfib, and Udo Zölzer, "Efficient Linear Prediction for Digital Audio Effects," in *Proc. Int. Conf. on Digital Audio Effects (DAFx'00)*, Verona, Italy, Dec. 2000.

[9] Mathieu Lagrange, Sylvain Marchand, Martin Raspaud, and Jean-Bernard Rault, "Enhanced Partial Tracking using Linear Prediction," in *Proc. Int. Conf. on Digital Audio Effects (DAFx'03)*, London, UK, Sep. 2003.

[10] Florian Keiler, Can Karadogan, Udo Zölzer, and Albrecht Schneider, "Analysis of Transient Musical Sounds by Auto-Regressive Modeling," in *Proc. Int. Conf. on Digital Audio Effects (DAFx'03)*, London, UK, Sep. 2003.

[11] Ismo Kauppinen, "Methods for Detecting Impulsive Noise in Speech and Audio Signals," in *Proc. 14th IEEE Int. Conf. on Digital Signal Processing (DSP2002)*, Santorini, Greece, Jul. 2002, pp. 963–966.

[12] Paulo Esquef, Matti Karjalainen, and Vesa Välimäki, "Detection of Clicks in Audio Signals Using Warped Linear Prediction," in *Proc. 14th IEEE Int. Conf. on Digital Signal Processing (DSP2002)*, Santorini, Greece, Jul. 2002, vol. 2, pp. 1085–1088.