ESTIMATION OF TIME-VARYING AR MODELS OF SPEECH THROUGH GAUSS-MARKOV MODELING

Krishna M Malladi

Hellosoft Inc.,
2099, Gateway Place, Suite #200,
San Jose, CA 95110
mohan@hellosoft.com

Ratnam V Rajakumar

Dept. of E&ECE,
Indian Institute of Technology,
Kharagpur, 721302, INDIA
rkumar@ece.iitkgp.ernet.in

ABSTRACT

In this work, a new method for estimating the time- varying AR model of speech is presented. Here, the time- varying parameters are modeled as stationary processes. Both the time- varying parameters and their corresponding stationary process are modeled through a common Gauss- Markov model whose state- vector can be estimated through the extended Kalman Filter (EKF) algorithm. The proposed algorithm is different from the earlier methods which use the EKF algorithm. Simulation studies are carried out for both voiced and unvoiced speech. It is shown that the proposed method has less mean-square prediction error than that obtained through the LPC method.

1. INTRODUCTION

Speech is typically a non-stationary signal [1] due to the change in its statistical characteristics with time. Model based spe- ech processing methods find extensive use in speech analysis as well as synthesis. Analysis and synthesis of speech has been a popular topic of research with many applications such as speech coding, speaker recognition, text to speech and speech to text conversion etc.

The LPC technique [2, 3] is the most popular technique used for the analysis and synthesis of speech signals. In this technique, the speech is windowed. The duration of the window is small and the windowed speech signal is represented by an all-pole model. The order of the all-pole model is typically 10. The LPC technique is applied on the all-pole model to extract the coefficients of the model. These coefficients are used to transmit and synthesize the speech.

The LPC technique divides the speech signal in to multiple frames and assumes quasi-stationary within the frame interval. But this technique is not useful to track the time-varying coefficients. The LPC technique smoothens any variations in the coefficients within the frame. Therefore, there is a need of a parametric technique which can track fast varying coefficients.

The Kalman filter is efficient in tracking the time-varying coefficients [4]. It is assumed in the literature that the time-varying parameter has a random walk which is not true. The time-samples of the parameter are correlated and are not random except for ocassional jumps. Therefore, the parameter walk can be modeled as stochastic process.

Earlier we developed a method for modeling and estimation of non- stationary signals by time-varying models such as time-varying AR models [5]. The time-varying coefficients are modeled as stationary AR processes. The time-varying AR model of the non- stationary signal and the parameter models are formed in to a Gauss- Markov model and the unified model is estimated through the extended Kalman filter (EKF). It is shown that the proposed method can estimate and track the fast-varying parameters.

There exists similar works in the literature but the present work is different from the earlier works. In the work of Iltis [6], though the time-varying parameters are modeled as stationary processes and the EKF algorithm is used for estimation, the parameters of those stationary models are assumed to be known and are not estimated. In the work of Tsatsanis et.al. [7], the parameters of the stationary models are estimated through least-squares method and the time-varying parameters are estimated through Kalman Filter. In the present work, the proposed method estimates both the time-varying parameters and their stationary models simultaneously using the EKF algorithm.

In the present work, speech is modeled as a time-varying AR process and the time-varying parameters are estimated through the above proposed method. Simulation results for both the voiced and unvoiced speech are presented. The performances of the LPC method and the proposed method are evaluated by comparing the prediction errors of both the methods.

1.1. Modeling the Speech

The speech $y(k)$ is modeled as a time-varying AR model of order p. The time-varying AR model can be expressed as

0-7803-7663-3/03/$17.00 ©2003 IEEE

[8],

$$y(k) = -\sum_{l=1}^{p} \theta_l(k) y(k-l) + v(k) \qquad (1)$$

where $\theta_l(k)$'s are the time-varying parameters of the model and $v(k)$ is the white Gaussian noise with variance σ_v^2.

1.2. Modeling the Parameter Walk

The time-varying parameters can be further modeled as a stochastic process [5]. The time-variation or the evolution of the parameters of a physical linear time-varying system is correlated. The variation may be considered to be largely deterministic with some random component. Therefore, the evolution process of the parameters can be modeled by a time-invariant AR process. Assuming that the parameter $\theta_l(k)$ of (1) can be modeled as an AR process of order r, the parameter walk model can be expressed as,

$$\theta_l(k) = -\sum_{s=1}^{r} a_{l_s} \theta_l(k-s) + w_l(k) \qquad (2)$$

where a_{l_s}'s are the fixed parameters of the stochastic model corresponding to $\theta_l(k)$ and $w_l(k)$ is the white Gaussian noise with the variance $\sigma_{w_l}^2$. The $w_l(k)$ is uncorrelated to $v(k)$.

2. THE GAUSS-MARKOV MODEL FORMULATION

In many situations, the stochastic model of the time-varying coefficients is not available, a priori. The parameters of the stochastic model has to be estimated simultaneously with the time-varying coefficients. In this work, we expressed (1) and (2) in the Gauss-Markov State-Space model form. This model will facilitate the estimation of the stochastic model parameters of the parameter walk and the time-varying coefficients using the EKF. The model formulation, State-Space model and the partial derivatives for the EKF algorithm are given in the following paragraphs.

From (1) and (2), the model of the non-stationary signal $y(t)$ can be expressed as

$$\Theta(k+1) = -\Phi_2(k) A(k) + W(k) \qquad (3)$$

$$y(k) = -\Phi_1(k)\Theta(k) + v(k) \qquad (4)$$

where,

$$\Theta(k) = [\theta_1(k)\, \theta_2(k)\, \ldots\, \theta_p(k)]^T$$

$$\Phi_2(k) = \begin{bmatrix} \Phi_{21}(k) & 0 & \ldots \\ & \Phi_{22}(k) & \ldots \\ & \ldots & \\ 0 & \ldots & \Phi_{2p}(k) \end{bmatrix}$$

$$\Phi_{2m}(k) = [\theta_m(k)\, \theta_m(k-1)\, \ldots\, \theta_m(k-r+1)]$$

$$A(k) = [a_{11}\, a_{12}\, \ldots\, a_{1r}\, \ldots\, a_{pr}]$$

$$W(k) = [w_1(k)\, w_2(k)\, \ldots w_r(k)]^T$$

$$\Phi_1(k) = [y(k-1)\, y(k-2) \ldots y(k-p)]$$

The equations (3)-(4) form the message model and the $\Theta(k)$ is the time-varying parameter vector. $\Phi_2(k)$ is of dimension (pxpr) and represents the regression matrix of $\Theta(k)$. $\Phi_{2m}(k)$ represents the regression vector of a time varying parameter $\theta_m(k)$ and $A(k)$ represents the time-varying parameters of the stochastic model of $\Theta(k)$.

In the above model formulation, the unknown parameters include the fixed $A(k)$ and time-varying $\Theta(k)$. The State-Space model form can be obtained by considering the augmented state-vector which can be expressed as,

$$X(k+1) = \begin{bmatrix} \Theta(k+1) \\ A(k+1) \end{bmatrix} \qquad (5)$$

In view of (3) and (5), the state updating equation can be obtained as

$$X(k+1) = \begin{bmatrix} -\Phi_2(k)A(k) \\ A(k+1) \end{bmatrix} + \begin{bmatrix} W(k) \end{bmatrix}$$

$$= \Phi\left[X(k), k\right] + W'(k) \qquad (6)$$

The observation model (4) can be written as

$$y(k) = h\left[X(k), k\right] + v(k) \qquad (7)$$

where,

$$h\left[X(k), k\right] = -\Phi_1(k)\Theta(k) \qquad (8)$$

Equations (6) and (7) constitute the State-Space model which involves nonlinear functions of the state-vector. With Gaussian excitation, it becomes the Gauss-Markov model.

3. MODEL ESTIMATION USING EKF

The Gauss-Markov model, formulated in the previous section, can be now used for the estimation of parameters through EKF. The EKF algorithm involves computation of partial derivatives from the Gauss-Markov model. Using the proposed model, the partial derivatives required in the EKF algorithm can be obtained . These are given by,

$$\frac{\partial h\left[\hat{X}(k+1/k), k+1\right]}{\partial \hat{X}(k+1/k)} = \frac{\partial \hat{y}(k+1)}{\partial \begin{bmatrix} \hat{\Theta}(k+1) \\ \hat{A}(k+1) \end{bmatrix}} \qquad (9)$$

and

$$\frac{\partial \Phi\left[\hat{X}(k/k), k\right]}{\partial \hat{X}(k/k)} = \frac{\partial \begin{bmatrix} \hat{\Theta}(k+1) \\ \hat{A}(k+1) \end{bmatrix}}{\partial \begin{bmatrix} \hat{\Theta}(k) \\ \hat{A}(k) \end{bmatrix}} \qquad (10)$$

The partial differentiations that are useful in evaluating (9) are given below in the scalar form.

$$\frac{\partial \hat{y}(k)}{\partial \hat{\theta}_j(k)} = -y(k-j) \qquad (11)$$

$$\frac{\partial \hat{y}(k)}{\partial \hat{a}_{ij}} = -\frac{\partial \hat{\theta}_i(k)}{\partial \hat{a}_{ij}} y(k-i) \quad (12)$$

$$\frac{\partial \hat{\theta}_i(k)}{\partial \hat{a}_{ij}} = -\hat{\theta}_i(k-j) - \sum_{t=1}^{r} a_{it} \frac{\partial \hat{\theta}_i(k-t)}{\partial \hat{a}_{ij}} \quad (13)$$

$$(14)$$

Similarly, the partial differentiations that are useful in (10) are given below.

$$\frac{\partial \hat{\theta}_l(k+1)}{\partial \hat{\theta}_l(k)} = -a_{1l1} \quad (15)$$

where a_{1l1} is the coefficient of time-varying parameter model $\theta_l(k)$ whose model order is reduced to one. Also,

$$\frac{\partial \hat{\theta}_l(k+1)}{\partial \hat{a}_{li}} = -\theta_l(k+1-i) - \sum_{t=1}^{r} a_{lt} \frac{\partial \hat{\theta}_l(k+1)}{\partial \hat{a}_{li}} \quad (16)$$

By evaluation and substitution of (9) and (10) in the EKF algorithm, the time-varying coefficients and their stochastic models can be simultaneously estimated. The EKF algorithm recursions are given by [9],

$$\begin{aligned} \hat{X}(k+1/k+1) &= \hat{X}(k/k-1) \\ &+ L_k \left[y(k) - h\left[\hat{X}(k-1/k), k\right]\right] \\ \hat{X}(k/k-1) &= \Phi\left[\hat{X}(k-1/k-1), k-1\right] \quad (17) \end{aligned}$$

The gain L_k is given by,

$$\Omega_k = H_k^T V_X(k/k-1) H_k + R_v \quad (18)$$

$$L_k = V_X(k/k-1) H_k \Omega_k^{-1} \quad (19)$$

where, V_X is the parameter error covariance, R_v is the covariance of $v(n)$, and

$$H_{k+1} = \frac{\partial h\left[\hat{X}(k+1/k), k+1\right]}{\partial \hat{X}(k+1/k)} \quad (20)$$

The updation equations for the error covariance is given by,

$$\begin{aligned} V_X(k/k) &= V_X(k/k-1) \\ &- V_X(k/k-1) H_k \Omega_k^{-1} H_k^T V_X(k/k-1) \\ V_X(k+1/k) &= F_k V_X(k/k) F_k^T + R_w \quad (21) \end{aligned}$$

where, R_w is the covariance of W' and

$$F_k = \frac{\partial \Phi\left[\hat{X}(k/k), k\right]}{\partial \hat{X}(k/k)} \quad (22)$$

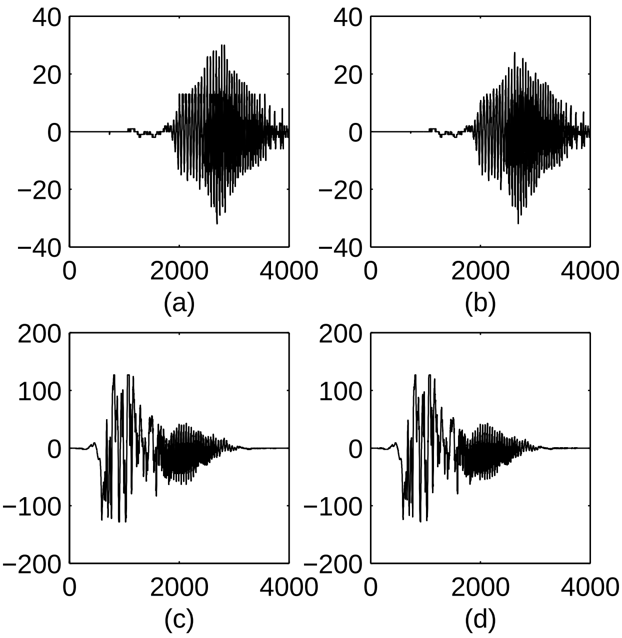

Fig. 1. The actual and the predicted signals (a) actual voiced speech (b) predicted voiced speech (c) actual unvoiced speech and (d) predicted unvoiced speech

4. SIMULATION STUDIES

Simulations are carried out to check the performance of the proposed unified algorithm and to compare with that of the LPC technique. Both voiced and unvoiced speech samples are considered for simulations.

A time-varying AR model of the order of 10 is considered to model the speech samples. Each time-varying parameter is modeled as a stationary AR process of order 2. The parameters for both the time-varying and the stationary models are estimated through the proposed unified model. The predicted speech samples are then obtained from the time-varying parameters.

Fig.1 shows the actual and the predicted speech samples. Fig. 1(a) shows the voiced speech. Fig.1(b) shows the predicted signal for the voiced speech. Fig.1(c) shows the unvoiced speech sample and Fig.1(d) shows the predicted speech for the unvoiced speech. It can be observed from the figures that the predicted speech samples are almost similar to the actual speech samples.

An AR model of order 10 is considered to estimate the time-varying parameters using the LPC technique. The same speech samples are used in the LPC technique. A new set of LPC parameters are estimated for every 200 samples. The predicted speech samples are then obtained from the estimated parameters of the AR model.

The prediction error is used to compare the performances of both the techniques. Mean-squared prediction errors are calculated for both the voiced and unvoiced signals and are shown below.

	voiced speech	unvoiced speech
signal power	66.5	826.25
mse of proposed method	1.056	7.1687
mse of LPC method	3.83	14.4419

It can be observed from the above that the proposed EKF technique has less prediction error as compared to LPC technique. The mean-square error is more in the case of unvoiced signal because the strength of the unvoiced signal is more than the voiced signal.

5. CONCLUSIONS

The proposed algorithm estimates and tracks the time- varying parameters of speech. The mean-square prediction error of the proposed technique is less than that of the LPC technique. Therefore, the proposed technique can be used for the analysis of speech.

6. REFERENCES

[1] Kondoz AM, Digital Speech coding for low bit rate communication systems, John Wiley and Sons, Chichester, 1994.

[2] Atal BS and Shroeder, “Adaptive predictive coding of speech signals,” the Bell system tech. jou., October 1970.

[3] Makhoul J, “Linear prediction: A tutorial review,” Proc. IEEE, pp561-580, April 1975.

[4] Ljung and Gunnarsson, ”Adaptation and tracking in system identification - A survey,” Automatica, vol.26, pp48-65, 1985.

[5] Krishna M Malladi, R.V.Raja Kumar and K. Veerabhadra Rao, ”A Gauss-Markov Model for the Estimation of ARMA Model of Time-varying signals and systems,” Intl. Syp. on Time-Frequency and Time-Scale Analysis, pp.657-660, Pittsburgh, October 7-9, 1998.

[6] R.A.Iltis, ”Joint estimation of PN code delay and multipath using extended Kalman Filter,” IEEE trans. on Communications, vol.38, no.10, pp.1677-1685, October 1990.

[7] M.K.Tsatsanis, G.B.Giannakis, G.Zhou, ”Estimation and equalization of fading channels with random coefficients,” Proc. ICASSP-96, pp.1093-1096, 1996.

[8] Y.Grenier,”Time-dependent ARMA modeling of nonstationary signals,” IEEE Trans. acoust., speech, signal proc., vol.31, No.4, pp 899-911, Aug.1983.

[9] B.D.O. Anderson and J.B. Moore, Optimal Filtering, Prentice Hall, 1979.