State Encoding of Hidden Markov Linear Prediction Models

Vikram Krishnamurthy and H. Vincent Poor

Abstract: **In this paper, we derive finite-dimensional non-linear filters for optimally reconstructing speech signals in Switched Prediction vocoders, Code Excited Linear Prediction (CELP) and Differential Pulse Code Modulation (DPCM). Our filter is an extension of the Hidden Markov filter.**

Index Terms: **Non-linear filtering, finite dimensional filters, state estimation, speech synthesis, Markov jump linear systems.**

I. INTRODUCTION AND PROBLEM FORMULATION

There are two main categories of speech coding algorithms [1]: *waveform coders* and *vocoders*. In waveform coders, the data transmitted from the encoder to decoder specify a representation of the original speech signal as a temporal waveform, so that the reproduced signal at the decoder approximates the original waveform. In contrast, vocoders do not reproduce an approximation to the original waveform; instead parameters that characterize individual sound segments are specified and transmitted to the decoder, which then constructs a new and different waveform. Vocoders operate at lower bit rates than waveform coders but the reproduced speech quality, while intelligible, usually suffers from the loss of naturalness and some of the unique characteristics of an individual speaker are often lost [1].

In this paper we derive a finite-dimensional non-linear filter that can be used in waveform coders as well as vocoders. Our filter is an extension of the Hidden Markov filter and optimally reconstructs the speech signal from the transmitted data. We will focus on the application of our non-linear reconstruction filter to three widely used algorithms, namely:

- Switched Prediction Coding (a vocoding algorithm)
- Code Excited Linear Prediction (CELP) (a waveform coding algorithm)
- Differential Pulse Code Modulation (DPCM) (a waveform coding algorithm)

The most basic linear predictive coder (LPC) for speech signals, models the human vocal tract as an auto-regressive (AR) process described by the difference equation [8, p. 270]:

$$r_n = \sum_{i=1}^{p} \zeta_i r_{n-i} + F_n u_n + G_n \bar{p}_n, \quad n \in Z^+ = \{1, 2, \dots\}, \tag{1}$$

where u_n denotes white noise, $\bar{p}_n$ denotes a periodic sequence of impulses with "pitch" period T_p, F_n and G_n are gains that allow for voiced speech ($F_n = 0$) and unvoiced speech ($G_n = 0$). 2-dimensional versions of the above model are also used in the coding of images (see Chapter 6 in [9]).

In vocoders, the encoder transmits the pitch period T_p, the gains G_n, F_n and the estimated AR model parameters ζ_i. The aim of the decoder is to reconstruct the speech signal r_n from this transmitted information.

In CELP, a codebook of a number of excitation sequences is maintained at encoder and decoder. After the AR parameters ζ_i are fit to the speech signal, the encoder searches the codebook for the excitation signal that produces the best fit to the original speech signal when the fitted AR parameters are used. Then, in addition to the information sent by the vocoder, the CELP encoder sends the code signifying which of the codebook excitation sequences gave the best fit. The receiver knows the codebook, and aims to reconstruct r_n from this information.

In DPCM (see [9], Chapter 6) the actual (quantized) prediction error excitation sequence is transmitted, rather than the closest match in the code book. Given this prediction error sequence, the receiver aims to reconstruct r_n.

In all three coding algorithms, a short-coming of the basic model (1) is that speech signals are known to be stationary only for short intervals of time (e.g., order of 20-30 msec) so that the AR coefficients ζ_i are constant only over such short intervals. Motivated by this non-stationarity, in this paper we focus on the above three speech coders that allow for time-varying parameters.

Problem Formulation

Let X_n denote an S-state discrete-time Markov chain defined on a probability space $(\Omega, \mathcal{F}, P)$ with state space $\{e_1, e_2, \dots, e_S\}$ where e_i denotes the unit S-vector with 1 in the ith position. Denote the transition probabilities $a_{ji} = P(X_n = e_j | X_{n-1} = e_i)$ and A for the $S \times S$ matrix $[a_{ji}]$, $1 \le i, j \le S$. Note that $\sum_{j=1}^{S} a_{ji} = 1$ for $1 \le i \le S$.

1. *Switched Prediction Vocoder*: Switched prediction vocoders belong to the class of forward adaptation algorithms (APF) (see

Manuscript received April 16, 1999; approved for publication by Iickho Song, Division I Editor, September 14, 1999.

Vikram Krishnamurthy is with Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3052, Australia, e-mail: v.krishnamurthy@ee.mu.oz.au.

H. Vincent Poor is with Department of Electrical Engineering, Princeton University, Princeton, NJ 08544, USA, e-mail: poor@princeton.edu.

This research was supported in part by the U.S. Office of Naval Research under Grant N00014-94-1-0115, and grants from the Australian Research Council and ATERB.

1229-2370/99/$10.00

Section 6.5 in [9]). In APF, fixed segments of input values are buffered and then transmitted after the p AR coefficients are estimated for the buffered segment.

In switched prediction vocoders [9, p. 301], both the transmitter and receiver have a bank of S predictors and adaptations consist of switching to one of these predictors according to some procedure. Since all the S predictors are pre-determined, there is no need to estimate, quantize or transmit the AR (predictor) coefficients. Therefore, the side information transmitted is merely the index of the predictor and this is significantly less than in other APF schemes.

In this paper we model the switched predictor as a Markov modulated AR process where the AR coefficients jump change according to the realization of a finite state Markov chain. Such a Markov modulated process captures the piecewise constant behaviour of the AR coefficients in a switched predictor. The switched predictor encoder assumes that the speech signal $r_n \in \mathbf{R}$ is generated as

$$r_n = \sum_{i=1}^{p} \zeta_i(X_{n-1})\, r_{n-i} + \langle F, X_{n-1}\rangle\, u_n + \langle G, X_{n-1}\rangle\, \bar{p}_n,$$
$$u_n \sim \text{white } N(0, \sigma_u), \quad (2)$$

where $F, G \in \mathbf{R}^S$ are gain vectors that allow for voiced or unvoiced speech and $\langle \cdot, \cdot \rangle$ denotes the inner product in $\mathbf{R}^S$. The encoder transmits the Markov chain X_n through a noisy additive white Gaussian noise channel. Our aim is to design a decoder that optimally reconstructs the speech signal r_n from the noisy observations y_n of X_n assuming that the AR coefficients $\zeta_i(X_n)$, pitch period and gains are known.

2. *CELP*: Our signal model here is identical to (2) except that the encoder also transmits the code of the excitation sequence $\bar{u}_n$ (that best fits the speech signal) to the decoder. Since the receiver also knows the codebook, assuming noise free transmission, the excitation signal $\bar{u}_n$ is known at the decoder.

3. *DPCM*: The signal model is identical to (2) except that the encoder also transmits the exact prediction error sequence. So assuming noise-less transmission, the receiver knows u_n. Such a Markov modulated DPCM system lies in the class of medium-to-high-complexity DPCM systems [9, p. 252].

In a sense, the three coders described above span the range from no knowledge to full knowledge of the excitation sequence u_n: no knowledge (Switched Prediction Vocoder), partial knowledge (CELP), and complete knowledge (DPCM). Notice in all three coders that computing the filtered estimates of the state of the Markov chain X_n from the noisy data y_n is straightforward: it is obtained using the standard Hidden Markov filter (discrete-time Wonham filter) [2]. However, it is not straightforward how to compute the filtered estimate $\hat{r}_n$ of r_n. Clearly the filtered estimate of $\zeta(X_{n-1}\, r_{n-i})$ is not equal to $\zeta(\hat{X}_{n-1})\, \hat{r}_{n-i}$, although the latter quantity is the conventional reconstruction. In this paper we will use the reference probability method [4] together with martingale methods to derive a *finite-dimensional filter* for reconstructing the speech signal r_n.

The existence of such a finite dimensional filter is of independent interest. It is well known that if instead of y_n, noisy measurements of r_n were available, the optimal filter for estimating r_n is *not* finite dimensional [5].

Similar Markov modulated time series are also used in failure detection and econometrics (see [3] and the references therein). In the special case $\bar{p}_n = 0$, the model is used for image-enhanced target tracking [10].

This paper is organized as follows: In Section I, we detail our signal model and estimation objectives. In Section III, finite dimensional discrete-time filters and smoothers are presented which yield estimates of the state of the jump linear system. Also computational issues are discussed.

II. GENERALIZED SIGNAL MODEL AND ESTIMATION OBJECTIVES

Motivated by the Markov modulated AR process described in Section I, we consider the following general signal model which includes the Switched predictor vocoder and CELP as special cases:

$$s_n = C(X_{n-1})\, s_{n-1} + \langle F, X_{n-1}\rangle\, v_n + \langle G, X_{n-1}\rangle\, p_n. \quad (3)$$

Here $X_n \in \{e_1, \ldots, e_s\}$ is the finite state Markov chain defined in Section I and s_n is assumed to be an N dimensional vector. Also for each state e_i of the Markov chain X_k, we assume that $C(e_i)$ is a $N \times N$ matrix. v_n is assumed to be a zero mean N-vector noise process (specified below) independent of X_n and p_n is a known N dimensional vector signal.

Remark: The jump linear model (3) includes the LPC model (2) as a special case. For example, the Markov modulated AR(2) process (with Markov modulated coefficients $\zeta_1(X_{n-1})$ and $\zeta_2(X_{n-1})$):

$$\begin{aligned} r_n &= \zeta_1(X_{n-1})\, r_{n-1} + \zeta_2(X_{n-1})\, r_{n-2} + F(X_{n-1})\, u_n \\ &\quad + G(X_{n-1})\, p_n, \qquad u_n \sim \text{white } N(0, \Sigma_u) \end{aligned} \quad (4)$$

can be modelled as (3) with $N = 2$,

$$s_n = \begin{bmatrix} r_n \\ r_{n-1} \end{bmatrix},$$
$$C(X_{n-1}) = \begin{bmatrix} \zeta_1(X_{n-1}) & \zeta_2(X_{n-1}) \\ 1 & 0 \end{bmatrix}, \quad (5)$$
$$v_n = \begin{bmatrix} u_n \\ 0 \end{bmatrix}.$$

□

- Switched Predictor Vocoder

Assume that X_n is transmitted through a noisy channel so that the decoder receives measurements $y_n \in \mathbf{R}$ where

$$y_n = \langle g, X_{n-1}\rangle + w_n, \quad w_n \sim \text{white } N(0, \Sigma_w). \quad (6)$$

In (6), $g = (g_1\ g_2 \ldots g_S)'$ is the vector of levels (drift-coefficients) of the Markov chain and $\langle \cdot, \cdot \rangle$ denotes the scalar product in $\mathbf{R}^S$. Also v_n is white noise unknown to the receiver.

- CELP

Apart from the Markov chain state X_n, the codeword μ_n for the excitation sequence $\bar{v}_n(\mu_n) \in \mathbf{R}^N$ that best fits the speech signal is also transmitted. We assume that this "best fit" is reflected as follows: Denoting the sigma algebra $\bar{\mathcal{V}}_n = \sigma\{\bar{v}_m, m \leq n\}$, we assume that $\mathbf{E}\{v_n(\mu_n)|\bar{\mathcal{V}}_n\} = \bar{v}_n(\mu_n)$, i.e., $\bar{v}_n(\mu_n)$ is the best estimate of v_n.

Since the set of possible codewords is finite, for convenience, we assume that the codeword process $\{\mu_n\}$, $n \in Z^+$, is modelled as a finite state Markov chain independent of the process $\{X_n\}$.

The CELP decoder receives two sets of observations: $y_n \in \mathbf{R}$ defined in (6) and z_n where

$$z_n = \mu_n + \eta_n, \qquad \eta_n \sim \text{ white } N(0, \Sigma_\eta).$$

We assume that w_n and η_n are mutually independent and also independent of X_n and v_n.

• DPCM

We assume that v_n is a correlated noise process because of the unmodelled dynamics caused by modelling errors such as selecting a low order AR model in (2). This information is transmitted through a noisy channel, so that the decoder receives two sets of measurements: $y_n \in \mathbf{R}$ (6) and $z_n \in \mathbf{R}^N$ where

$$z_n = v_n + \eta_n, \qquad \eta_n \sim \text{ white } N(0, \Sigma_\eta).$$

We assume that w_n and η_n are mutually independent and also independent of X_n and v_n.

Notation: For any $n \in Z^+$, define the following sigma-algebras: $\mathcal{F}_n = \sigma\{X_m, m \leq n\}$, $\mathcal{Y}_n = \sigma\{y_m, m \leq n\}$, $\mathcal{V}_n = \sigma\{v_m, m \leq n\}$, $\mathcal{Z}_n = \sigma\{z_m, m \leq n\}$, $\mathcal{G}_n = \mathcal{F}_{n-1} \bigvee \mathcal{Y}_n \bigvee \mathcal{Z}_n \bigvee \mathcal{V}_n$, i.e., the sigma field generated by $\{X_{m-1}, y_m, z_m, v_m\}$, $m \leq n$, $\mathcal{O}_n = \mathcal{Z}_n \bigvee \mathcal{Y}_n$.

For any measurable process $\{\phi_n\}$, $n \in Z^+$, let $\hat{\phi}_n = \mathbf{E}\{\phi_n|\mathcal{O}_n\}$, where $\mathbf{E}$ denotes expectation under the probability measure P.

Aim: For fixed known values of A, g and Σ_w, our aim is to design a filter at the decoder that optimally reconstructs the filtered estimates $\hat{s}_n = \mathbf{E}\{s_n|\mathcal{O}_n\}$.

Remarks: 1. We do *not* address parameter estimation in this paper. Indeed, given $\mathcal{O}_n$, the maximum likelihood estimates of A, g, Σ_w are readily obtained using the Expectation Maximization (EM) algorithm (Baum Welch equations), see [2] or [12] for details.

2. Note that in (6), the observation at time n depends on the Markov state at time $n-1$. This is reasonable from a dynamical point of view as the reaction due to X_{n-1} is not instantaneous. It is also possible to derive a filter when $y_n = \langle g, X_n\rangle + w_n$. We refer the reader to [6] for details. □

III. RECONSTRUCTION OF THE ENCODED SIGNAL

In this section, we derive an optimal finite dimensional filter for computing $\hat{s}_n$.

A. Preliminaries

It is straightforward to show that the semi-martingale representation of the Markov chain X_n is [6]

$$X_n = A X_{n-1} + M_n, \tag{7}$$

where M_n is a $(P, \mathcal{F}_n)$ martingale increment.

We shall use the reference probability method to derive our filters.

Define the probability measure P_0 such that the $\mathcal{G}_{n-1}$ restriction of the Radon-Nikodym derivative of P with respect to P_0 is

$$\begin{aligned} \left.\frac{dP}{dP_0}\right|_{\mathcal{G}_{n-1}} &= \Lambda_n = \prod_{m=1}^{n} \gamma_m, \\ \text{where } \gamma_m(X_{m-1}, y_m) &\\ &= \exp\left(-\frac{1}{2\,\Sigma_w}(\langle g, X_{m-1}\rangle^2 - 2\, y_m\, \langle g, X_{m-1}\rangle)\right). \end{aligned} \tag{8}$$

Then the following results hold [6]:

1. Λ_n is a (P_0, G_n) martingale.
2. Under P_0, y_n, $n \in Z^+$ is a $N(0, \Sigma_w)$ white process independent of X_n. (This is a discrete-time version of Girsanov's theorem).
3. If ϕ_n, $n \in Z^+$ is a measurable sequence, then an abstract version of Bayes theorem states

$$\hat{\phi}_n = \mathbf{E}\{\phi_n|\mathcal{O}_n\} = \frac{\mathbf{E}_0\{\Lambda_n\, \phi_n|\mathcal{O}_n\}}{\mathbf{E}_0\{\Lambda_n|\mathcal{O}_n\}}. \tag{9}$$

For notational convenience, define the *unnormalized conditional expectation* $\sigma_n(\phi_n) = \mathbf{E}_0\{\Lambda_n\, \phi_n|\mathcal{O}_n\}$. Then $\hat{\phi}_n$ in (9) can be re-expressed as

$$\begin{aligned} &\hat{\phi}_n = \sigma_n(\phi_n)/\sigma_n(1), \\ &\quad\text{where } \sigma_n(1) = \mathbf{E}_0\{\Lambda_n|\mathcal{O}_n\} = \mathbf{E}_0\{\Lambda_n|\mathcal{Y}_n\}, \end{aligned} \tag{10}$$

where the last equality follows because Λ_n and $\mathcal{Y}_n$ are independent of $\mathcal{Z}_n$.

B. Zakai State Filter

For notational convenience, let $b_i(y_m) = \gamma_m(e_i, y_m)$ where γ_m is defined in (8), that is:

$$b_i(y_m) = \exp\left(-\frac{1}{2\,\Sigma_w}(g_i^2 - 2\, y_m\, g_i)\right), \quad i = 1, \ldots, S. \tag{11}$$

In the following theorem, we derive a recursive filter for $\sigma_n(s_n)$.

Theorem 1: The filtered state is given by $\sigma_n(s_n) = \sum_{i=1}^{S} \sigma_n(s_n\, X_n(i))$ where

$$\begin{aligned} \sigma_n(s_n\, X_n(i)) &= \sum_{j=1}^{S} C(e_j)\, \sigma_{n-1}(s_{n-1}\, X_{n-1}(j)) a_{ij}\, b_j(y_n) \\ &\quad + [F(j)\, \sigma_n(v_n) + G(j)\, p_n]\, \sigma_{n-1}(X_{n-1}(j))\, a_{ij}\, b_j(y_n). \end{aligned} \tag{12}$$

Proof: Let $e_k \in \mathbf{R}^N$ denote the unit N-vector with 1 in the k-th position. Then the k-th component of s_n is

$$\begin{aligned} s_n(k) &= \langle s_n, e_k\rangle \\ &= e_k'\, C(X_{n-1})\, s_{n-1} + \langle F, X_{n-1}\rangle v_n(k) + \langle G, X_{n-1}\rangle p_n(k). \end{aligned} \tag{13}$$

From (7) and (13), we have

$$s_n(k)\,X_n = e_k'\,C(X_{n-1})\,s_{n-1}\,A\,X_{n-1} + [\langle F, X_{n-1}\rangle v_n(k) + \langle G, X_{n-1}\rangle p_n(k)]\,A\,X_{n-1} + \mathcal{G}_n \text{ martingale increment.} \quad (14)$$

Then multiplying by Λ_n and taking conditional expectations with respect to $\mathcal{O}_n$, we have

$$\sigma_n(s_n(k)\,X_n) = e_k'\,\sigma_n(C(X_{n-1})\,s_{n-1}\,A\,X_{n-1}) + \cdots . \quad (15)$$

However, this is not in a recursive form because of the first term on the RHS. So let us go back to (14). The trick is to take inner products with the unit vector e_i. So defining $X_t(i) = \langle X_t, e_i\rangle$, and post-multiplying both sides of (14) by e_i yields:

$$s_n(k)\,X_n(i) = e_k'\,C(X_{n-1})\,s_{n-1}\langle A\,X_{n-1}, e_i\rangle + [\langle F, X_{n-1}\rangle v_n(k) + \langle G, X_{n-1}\rangle p_n(k)]\ \langle A\,X_{n-1}, e_i\rangle + \mathcal{G}_n \text{ martingale increment.} \quad (16)$$

Now using the identity $\langle A\,X_{n-1}, e_i\rangle = \sum_{j=1}^{S}\langle A\,\langle X_{k-1}, e_j\rangle, e_i\rangle = \sum_{j=1}^{S}\langle X_{n-1}, e_j\rangle\,a_{ij}$, we have

$$s_n(k)\,X_n(i) = \sum_{j=1}^{S} e_k'\,C(e_j)\,s_{n-1}\,X_{n-1}(j)\,a_{ij} + [F(j)\,v_n(k) + G(j)\,p_n(k)]\,X_{n-1}(j)\,a_{ij} + \mathcal{G}_n \text{ martingale increment.} \quad (17)$$

Multiplying both sides of the above equation by Λ_n yields

$$\Lambda_n\,s_n(k)\,X_n(i) = \sum_{j=1}^{S} e_k'\,C(e_j)\,s_{n-1}\,X_{n-1}(j)\,a_{ij}\,\Lambda_{n-1}\,\gamma_n + \Lambda_{n-1}\,[F(j)\,v_n(k) + G(j)p_n(k)]\ X_{n-1}(j)\,a_{ij}\,\gamma_n + \mathcal{G}_n \text{ martingale increment.}$$

Using the fact that $X_{n-1}(j)\,\gamma_n = X_{n-1}(j)\,b_j(y_n)$ and taking conditional expectations with respect to $\mathcal{O}_n$ then yields

$$\sigma_n(s_n(k)\,X_n(i)) = \sum_{j=1}^{S} e_k'\,C(e_j)\,\sigma_{n-1}(s_{n-1}\,X_{n-1}(j))\,a_{ij}\,b_j(y_n) + [F(j)\,\sigma_n(v_n(k)) + G(j)\,p_n(k)]\,\sigma_{n-1}(X_{n-1}(j))\,a_{ij}\,b_j(y_n). \quad (18)$$

Finally stacking together $(\sigma_n(s_n(1)\,X_n(i)), \sigma_n(s_n(2)\,X_n(i)), \ldots, \sigma_n(s_n(N)\,X_n(i)))$ and noting that $(e_1, e_2, \ldots, e_s)$ is merely the $S \times S$ identity matrix yields the above result (12). □

To compute $\hat{s}_n$, we use (10) and Theorem 1:

$$\hat{s}_n = \sigma_n(s_n)/\sigma_n(1), \quad \text{where } \sigma_n(1) = \sum_{j=1}^{S}\sigma_n(X_n(j)). \quad (19)$$

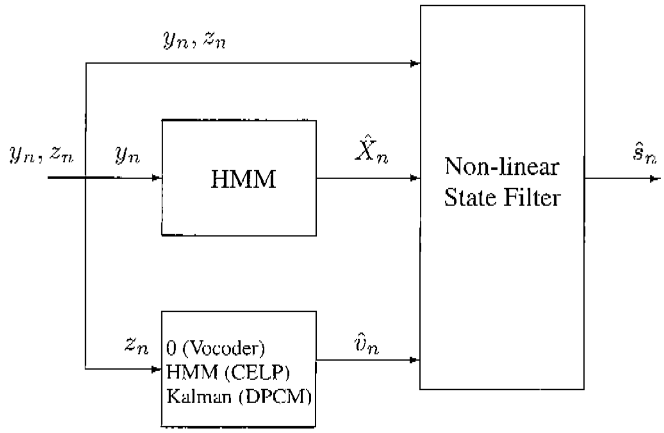


Fig. 1. Optimal Non-linear Decoder.

In the above equation, the unnormalized state estimate $\sigma_n(X_n(j))$ is computed using the standard HMM state filter [6], [2]

$$\sigma_n(X_n(i)) = \sum_{j=1}^{S}\sigma_n(X_{n-1}(j))\,a_{ij}\,b_j(y_n). \quad (20)$$

In summary, (12) and (20) together with the normalization (19) provide a recursive optimal filter for $\hat{s}_n$. It only remains to compute the filtered estimate $\sigma_n(v_n)$ for the three cases mentioned in Section II.

1. Switched Prediction Vocoder: In this case, v_n is white noise unknown to the decoder; therefore $\sigma_n(v_n) = \mathbf{E}\{v_n\} = 0$.

2. CELP: Since the code-book is known at the decoder, assuming noise free transmission of the coderword μ_n (i.e., $\eta_n = 0$), then $\sigma_n(v_n) = \bar{v}_n(\mu_n)$.
For non-zero channel noise (i.e., η_n is non-zero), $\hat{\mu}_n = \mathbf{E}\{\mu_n|\mathcal{Z}_n\}$ can be obtained via the standard Hidden Markov Model filter and $\sigma_n(v_n) = \bar{v}_n(\hat{\mu}_n)$. Of course $\hat{\mu}_n$ is continuous-valued whereas the set of codewords is finite. So in practice, the closest code vector in the codebook to $\hat{\mu}_n$ can be chosen. Alternatively, the Viterbi algorithm can be used to compute maximum likelihood sequence estimates of the code vectors.

3. DPCM: As mentioned in Section II, in this case, v_n is a correlated process transmitted via a noisy channel. In particular, if v_n is a correlated linear process, then $\sigma_n(v_n)$ is obtained from the Kalman filter (see [11] for the Kalman filter equations).

Fig. 1 shows the setup of our optimal decoder. In Case 1, $\hat{v}_n = 0$ and the decoder consists of the standard HMM filter (20) together with our non-linear filter (12). In Case 2, $\hat{v}_n$ is obtained as the output of another HMM filter. (If noise free transmission is assumed, then $v_n = \bar{v}_n$.) In Case 3, $\hat{v}_n$ is obtained as the output of a Kalman filter.

C. Smoothing

Our aim here is to compute the smoothed estimate $\mathbf{E}\{s_m(k)|\mathcal{O}_n\}$, $m < n$. We only consider the cases where v_n is either white noise or a known sequence to the decoder.

For $m < n$, we have

$$s_m(k)\, X_n = s_m(k)\, A\, X_{n-1} + \mathcal{G} \text{ martingale increment,} \tag{21}$$

So

$$\Lambda_n\, s_m(k)\, X_n = \Lambda_{n-1}\, A\, s_m(k)\, X_{n-1}\, \gamma_n + \mathcal{G} \text{ martingale increment.} \tag{22}$$

Taking inner product with e_i and because $X_{n-1} = \sum_{j=1}^{S} X_{n-1}(j)\, e_j$, we have the following result.

Corollary 1: The smoothed estimate $\mathbf{E}\{s_m(k)|\mathcal{O}_n\}$ is computed as

$$\mathbf{E}\{s_m(k)|\mathcal{O}_n\} = \frac{\sigma_n(s_m(k))}{\sigma_n(1)} \quad \text{where } \sigma_n(s_m(k)) = \sum_{i=1}^{S} \sigma_n(s_m(k)\, X_n(i)), \tag{23}$$

$$\sigma_n(s_m(k)\, X_n(i)) = \sum_{j=1}^{S} a_{ij}\, \sigma_{n-1}(s_m(k)\, X_{n-1}(j))\, b_j(y_n). \tag{24}$$

D. Computational Issues

The complexity (number of real multiplications) for computing the Zakai filter $\sigma_n(s_n\, X_n(i))$ is $O(S^2)$ for each i, $i = 1, \ldots, S$. So the complexity required in computing $\sigma_n(s_n)$ is $O(S^3)$ at each time instant.

Also the computational complexity involved in computing the HMM state filter $\sigma_n(X_n)$ (20) is $O(S^2)$.

The filters derived above have a similar structure to the Hidden Markov Model filters derived in [7]. Therefore the same techniques used in [7] can be used for normalizing and scaling the filtered variables to avoid numerical underflow. Also as in [7], our filters are suitable for parallel implementation e.g., on a systolic array architecture. In particular, note that the filter for $\sigma_n(s_n\, X_n(i))$ is independent of $\sigma_n(s_n\, X_n(i'))$, $i \neq i'$. So $\sigma_n(s_n\, X_n(i))$ can be computed in parallel for each i, $i = 1, \ldots, S$.

IV. DISCUSSION

Evaluating the performance of a speech coding algorithm is rather subjective since it requires human listeners to judge the intelligibility of the reconstructed speech signal. Therefore, we stress that although our filter optimally reconstructs the speech signal (under the modelling assumptions), practical studies are required to evaluate the performance of the algorithm on actual speech.

REFERENCES

[1] A. Gersho, "Advances in speech and audio compression," *Proc. IEEE*, vol. 82, no. 6, pp. 900–918, June 1994.

[2] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. IEEE*, vol. 77, no. 2, pp. 257–285, 1989.

[3] V. Krishnamurthy and T. Ryden, "Consistent estimation of linear and nonlinear autoregressive models with Markov regime," *J. Time Series Analysis*, vol. 19, no. 3, pp. 291–308, May 1998.

[4] R. J. Elliott, *Stochastic Calculus and Applications*, Springer Verlag, NY, 1982.

[5] J. Tugnait, "Adaptive estimation and identification for discrete systems with Markov jump parameters," *IEEE Trans. Auto. Control*, vol. AC-27, no. 5, Oct. 1982.

[6] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models – Estimation and Control*, Springer Verlag, 1995.

[7] M. James, V. Krishnamurthy and F. LeGland, "Time discretization of continuous-time filters and smoothers for HMM parameter estimation," *IEEE Trans. Inform. Theory*, vol. 32, no. 2, pp. 593–605, Mar. 1996.

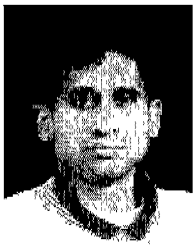
[8] J. G. Proakis and M. Salehi, *Communication Systems Engineering*, Prentice Hall, NJ, 1994.

[9] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, NJ, 1984.

[10] V. Krishnamurthy and J. S. Evans, "Covert Tracking of Markov Jump Linear Systems," *Automatica*, vol. 34, no. 6, pp. 765–770, June 1998.

[11] B. D. O. Anderson and J. B. Moore, *Optimal filtering*, Prentice Hall, Englewood Cliffs, New Jersey, 1979.

[12] L. Deng and C. Rathinavelu, "A Markov model containing state-conditioned second-order non-stationarity: Application to speech recognition," *Computer Speech and Language*, no. 9, p. 63–86, 1995.



Vikram Krishnamurthy was born in 1966. He received his bachelor's degree in Electrical Engineering from the University of Auckland, New Zealand in 1988, and doctoral degree from the Australian National University, Canberra in 1992. He is currently an Associate Professor and Deputy Head of the Department of Electrical Engineering, University of Melbourne, Australia. His research interests are in time-series analysis, stochastic filtering theory, and statistical signal processing in communication systems. Dr Krishnamurthy is currently an associate editor for *IEEE Transactions on Signal Processing* and *Systems and Control Letters*.

H. Vincent Poor is Professor of Electrical Engineering at Princeton University. He received the Ph.D. degree in Electrical Engineering and Computer Science from Princeton in 1977. From 1977 until he joined the Princeton faculty in 1990, he was a faculty member at the University of Illinois. He has also held visiting and summer appointments at several universities and research organizations in the U.S., Britain and Australia. Dr. Poor's research interests are in the area of statistical signal processing and its applications, primarily in wireless communications. His publications in this area include the the recent book, *Wireless Communications: Signal Processing Perspectives*, co-edited with Gregory Wornell.

Dr. Poor is a Fellow of the IEEE, the Acoustical Society of America, and the American Association for the Advancement of Science. He has been involved in a number of IEEE activities, including serving as President of the IEEE Information Theory Society in 1990, and as a member of the IEEE Board of Directors in 1991 and 1992. He recently served as Technical Co-chair for the "Golden Jubilee" IEEE International Symposium on Information Theory, held at MIT in August 1998. Among his awards are the 1992 Terman Award from the American Society for Engineering Education, and the 1994 Distinguished Member Award of the IEEE Control Systems Society.