A COMPARATIVE PERFORMANCE EVALUATION OF ADAPTIVE ARMA SPECTRAL ESTIMATION METHODS FOR NOISY SPEECH

ANJAN BASU AND K. K. PALIWAL

COMPUTER SYSTEMS AND COMMUNICATIONS GROUP
TATA INSTITUTE OF FUNDAMENTAL RESEARCH
HOMI BHABHA ROAD, COLABA
BOMBAY 400 005
INDIA

Abstract: In this paper, the problem of adaptive estimation of linear prediction (LP) coefficients from noisy speech is considered. Performance of the following three adaptive ARMA spectral estimation algorithms are studied for this purpose: 1) the Recursive extended least squares (RELS) algorithm, 2) the Recursive maximum likelihood (RML) algorithm and 3) the Overdetermined recursive instrumental variable (ORIV) algorithm. To put them in proper perspective, the normalized LMS (NLMS) has also been considered. The ORIV algorithm is found to be the best in terms of Itakura distance from the ideal LP coefficients and the power spectral density estimation. The RML algorithm is found to be robust in highly noisy cases.

I. INTRODUCTION

Linear prediction analysis of speech has played a major role in speech processing applications since its first introduction by Saito and Itakura in 1966 [1] and Atal and Schroeder in 1967 [2]. Success of LP analysis can be attributed to the fact that speech characteristics are more or less determined by the positions of its spectral peaks or formants and this type of *peaky* behaviour of the spectrum can be nicely represented by an autoregressive (AR) model; *i.e.*, an all-pole model excited by white Gaussian noise. Many efficient techniques exist for the estimation of LP parameters of speech [3]. But addition of white noise to speech makes this problem difficult, because noisy speech can no longer be represented by an AR process. In fact, it can be shown that addition of white noise to speech makes it an autoregressive moving-average (ARMA) process. Because of this, conventional LP analysis methods perform very poorly for noisy speech. In addition to this noisy case, there are some speech sounds like nasals where spectral peaks as well as vallies are important and, hence, a pole-zero model or an ARMA representation is necessary for them. Therefore, it is necessary to use ARMA modelling for accurate estimation of LP parameters of speech in these cases.

Most of the speech applications require estimation of LP parameters in a block mode; *i.e.*, the speech signal is divided into a number of segments and the segments are processed independently. However, there are some applications (such as ADPCM type of speech coding) where it is necessary to process speech in an adaptive (or, sequential) mode for sample-by-sample estimation of LP parameters. The present paper addresses this adaptive LP parameter estimation problem for noisy speech.

In the present paper, we explore different adaptive ARMA spectral estimation methods for LP parameter estimation of noisy speech. We consider the following three algorithms for this purpose: 1) Recursive extended least squares (RELS) algorithm [4], 2) Recursive maximum likelihood (RML) algorithm [4,5], and 3) Overdetermined recursive instrumental variable (ORIV) algorithm [6]. In order to put these algorithms in proper perspective, we also compare their performance with the popular normalized LMS (NLMS) algorithm [7].

II. MATHEMATICAL FORMULATION

Let us consider the speech signal $x(t)$ which can be modeled by an AR process of order p

$$x(t) = \sum_{i=1}^{p} a_i\, x(t-i) + u(t) \qquad (1)$$

where $u(t)$ is the excitation signal having flat spectrum. Speech signal $x(t)$ gets corrupted by an additive white noise $n(t)$ and we observe only the noisy speech

CH2561-9/88/0000-0675 $1.00 © 1988 IEEE

$y(t)$. Our aim here is to compute the sample-by-sample estimate of LP coefficients $\{a_i\}$ from the noisy speech signal $y(t)$. In order to do it, we observe that the noisy speech signal can be represented as an ARMA model of order (p,p) and it follows the difference equation

$$y(t) = \sum_{i=1}^{p} a_i y(t-i) + \sum_{i=0}^{p} b_i \xi(t-i) \qquad (2)$$

where the uncorrelated process $\xi(t)$ depends on $n(t)$ and $u(t)$. Here we note that the coefficients of AR part of the ARMA model are identical to the required LP coefficients of the speech signal. Therefore, the conventional adaptive ARMA methods can be used for estimating the LP parameters from noisy speech.

III. ADAPTIVE ARMA ALGORITHMS

Let us rewrite the equation (2) for time varying case

$$y(t) = \theta^T(t)\ y(t-1) + \xi(t) \qquad (3)$$

where $\theta(t)$ and $y(t-1)$ are parameter vector and data vector, respectively,

$$\theta^T(t) = [a_1(t), \ldots, a_p(t), b_1(t), \ldots, b_p],$$

$$y(t-1) = [y(t-1), \ldots, y(t-p), \xi(t-1), \ldots, \xi(t-p)].$$

An adaptive ARMA estimation algorithm estimates the parameter vector $\theta(t)$ at each time step t from a finite set of data points upto t, $\mathbf{Y}_t = \{y(1), y(2), \ldots, y(t)\}$. To tackle the time variation of parameters, the data points are exponentially weighted as follows.

$$\bar{y}(i) = \lambda^{t-i}\ y(i), \quad 0 < \lambda < 1$$

whete λ is the forgetting factor. Thus, the weightage given to the data points in the past are gradually reduced. For faster convergence of algorithms, λ is kept small at the beginning and as the time progresses it keeps increasing and comes to a steady value close to one. A typical method of varying λ with time is given below.

$$\lambda_t = \begin{cases} \alpha\, \lambda_{t-1} + (1-\alpha), & \text{for } \lambda_t < \lambda^o \\ \lambda^o, & \text{for } \lambda_t \geq \lambda^o \end{cases} \qquad (4)$$

In this paper, we have taken $\lambda_o = 0.95$ and $\alpha = 0.99$. λ^o is taken to be less than unity, but its actual value is different for different adaptive ARMA algorithms.

With these preliminaries, we describe below the different adaptive ARMA algorithms.

Recursive Maximum Likelihood Algorithm: The RML algorithm is based on the approximate maximization of the likelihood function of observed data sequence, $\mathbf{Y}_t = \{y(1), y(2), \ldots, y(t)\}$, parameterized by ARMA coefficient vector. The details of its derivation can be found elsewhere [7]. Only a summary of the algorithm is presented below.

Let $\hat{\theta}(t)$ be the estimate of $\theta(t)$ using data points upto time t and is given by:

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t)\ \psi(t-1)\ \epsilon(t) \qquad (5)$$

where $P(t)$ is the error covariance matrix and is given by the following recursion:

$$P(t) = [P(t-1) - P(t-1)\psi(t-1)\psi^T(t-1)P(t-1) / (\lambda_t - \psi^T(t-1)P(t-1)\psi(t-1))] / \lambda_t + [\lambda_t + \psi^T(t-1)P(t-1)\psi(t-1)] \qquad (6)$$

Here, the vector $\psi(t-1)$ is the 'pre-filtered' data vector and will be defined later. The other quantities are defined below.

$$\epsilon(t) = \text{a priori error}$$

$$= y(t) - \hat{\theta}^T(t-1)\ \phi(t-1)$$

$$\phi(t-1) = \text{approximate data vector}$$

$$= [y(t-1), \ldots, y(t-p), e(t-1), \ldots, e(t-p)]$$

$$e(t) = \text{a posteriori error (residual error)}$$

$$= y(t) - \hat{\theta}^T(t)\phi(t-1)$$

The filtered data vector $\psi(t-1)$ is obtained by passing $\phi(t-1)$ through a pre-filter $1/\hat{B}(kz^{-1})$. The pre-filter is obtained from the MA part of the estmated parameter vector $\theta(t)$ by the following relation.

$$\hat{B}(kz^{-1}) = 1 + k\hat{b}_1(t)z^{-1} + \ldots + k^p\hat{b}_p(t)z^{-p}$$

For actual RML algorithm, k=1. But the disadvantage associated with it is that continuous monitoring of the stability of the 'pre-filter' is needed. Moreover for a narrowband signal (like speech), where roots are close to the unit circle, the system response becomes very slow; hence the algorithm takes a long time to converge. Friedlander [8] suggested a 'modified pre-filter' with value of k < 1. This has the effect of pulling the roots towards the centre of the unit circle. It reduces the time constant of the system and the convergence becomes faster.

Recursive Extended Least Squares Algorithm: RELS algorithm is actually a pseudolinear regression method. In this method unobservable innovation sequance is approximated by the residual error sequence, and commonly known recursive least-squares algorithm is applied on it. In the RML algorithm, if the pre-filter parameter 'k' is set to zero (*i.e.*, no pre filtering is done), it becomes the RELS algorithm. It can be shown that this algorithm does not guarantee convergenceto the actual parameter values [10]; but on the other hand it needs no stability monitoring and has faster convergence than the RML algorithm in the narrowband case.

Overdetermined Recursive Instrumental Variable Method: Instrumental variable (IV) method of ARMA process estimation provides consistent estimates of AR parameters, without estimating noise parameters. There can be various choices of the instrument (or the instrumental variable); one common choice is the delayed data vector $\mathbf{z}_t = \mathbf{y}_{t-q} = [y(t-q), \ldots, y(t-2q+1)]$, where q is the order of the MA part of the ARMA process. One method of improving the accuracy of estimation is to use more *instruments* than what is necessary and solve the overdetermined set of linear equations (k equations in p variables, k > p) in a least-squares sense. This method formulated in a recursive algorithm is the ORIV method. Details about this method can be found in [6].

IV. EXPERIMENTAL RESULTS

In order to evaluate the performance of different ARMA algorithms, we use a long sequence of speech data (male voice) sampled at 8 kHz. The noisy version of data is obtained by addition of Gaussian distributed pseudo-random numbers. For comparison of performances of different algorithms, it is necessary 1) to obtain the ideal LP parameters for each data point which can be used as a reference; and 2) to select a meaningful distance measure between the ideal LP parameters and the estimated parameters at each time step and find out the mean distance over a long data sequence.

The reference LP parameters are obtained from clean speech at each point using the autocorrelation method with a 16 ms sliding Hamming window. We have used log-likelihood or Itakura distance measure [11] for calculating the deviation of estimated LP parameters from the ideal ones. Fig. 1 shows the temporal variations of Itakura distance over 5000 data points for the following algorithms: 1) the NLMS algorithm, 2) the RELS algorithm, 3) the RML algorithm, 4) the ORIV algorithm with order of the over-determined system, k = 2p and 5) the ORIV algorithm with k = 3p. Plots are given for the 10 dB signal-to-noise ratio (SNR) case. The corresponding energy countour is also shown at the top. Comparision with the energy contour shows that these values shoot up (*i.e.*, error in estimation increases) at the time of change in energy level. The figure also shows that at the high energy portion Itakura distance is low for all algorithms; this happens because SNR at those portions is high.

Table 1 gives the average Itakura distance for the same five algorithms Results are given for 10 dB and 5 dB SNR cases. It can be seen from this table that only the ORIV algorithm gives good result, while the RELS and RML algorithms are even worse than the normalized LMS algorithm. This is because the RELS and RML algorithms use a posteriori error (or residual) as an estimate of the innovation sequence to the ARMA model. Since this is

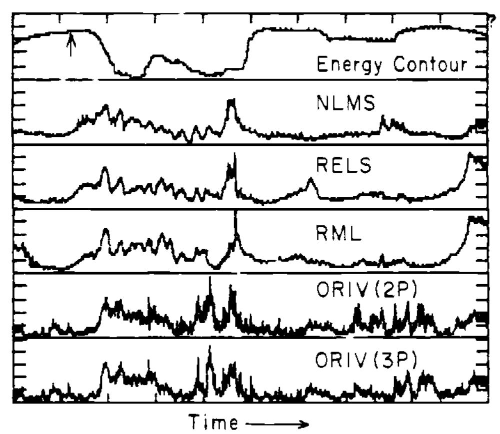

Fig. 1: Itakura distance measure as a function of time for different ARMA algorithms.

Table 1.

Algorithms	Itakura distance	
	10 dB	5 dB
1. NLMS	1.28	1.53
2. RELS	1.35	1.52
3. RML	1.53	1.56
4. ORIV (k=2p)	1.11	1.48
5. ORIV (k=3p)	1.05	1.32

not a good estimate in case of noisy speech, the RELS and the RML algorithms show higher Itakura distance. On the other hand the ORIV method does not need this estimate and, thus, it is not prone to this error. In addition, it is observed that performance of ORIV algorithm improves with an increase in the order of the over-determined system. It has also been observed that increase in order beyond some point starts degrading the estimate. The RML algorithm gives high Itakura distance; but it can be observed that it is very much immune to the noise, so the distance measure does not increase significantly from 10 dB to 5 dB SNR case. In highly noisy cases the RML algorithm becomes better than the others.

In order to compare the spectral estimation performance of different ARMA algorithms, we compute the power spectra at a particular time instant (marked by an arrow in the Fig. 1) using different algorithms. Results are shown in Fig. 2. The ideal power spectrum estimated from the clean speech is also shown in the figure for reference. It can be seen from this figure that the ORIV algorithm (with k = 3p) results in the best power spectral estimation; other algorithms estimate the power spectrum with a spurious peak between the first two formants.

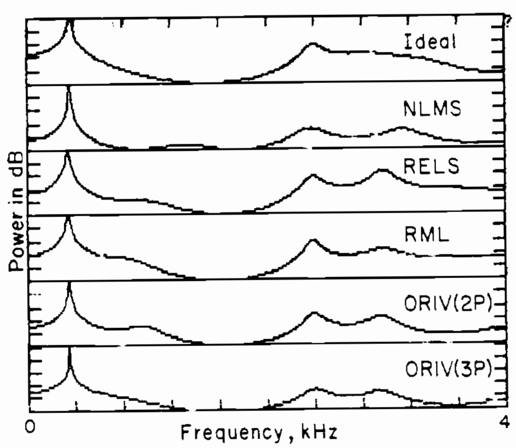

Fig. 2: Power spectral density estimates for different ARMA algorithms.

V. CONCLUSION

In this paper the problem of sample-by-sample estimation of LP coefficients from noisy speech is addressed. Different adaptive ARMA spectral estimation algorithms are studied for this purpose and their comparative performance evaluation is done. The ORIV algorithm (with k = 3p) is found to be the best algorithm. Though the RML algorithm is not as good as the ORIV algorithm for higher SNR, it is more immune to further addition of noise, *i.e.*, its performance does not degrade much with further increase in noise level.

References

[1] S. Saito and F. Itakura, "The theoritical consideration of statistically optimum method for speech spectral density", Rept. No. 3170, Elec. Comm. Lab, N.T.T., Tokyo, 1966.

[2] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals", Proc. 1967 Conf. Comm. and Process, 360-361, 1967.

[3] J. D. Markel and A. H. Gray, Jr., "Linear prediction of speech", Springer-Verlag, 1976.

[4] T. Soderstrom, L. Ljung and I. Gustavsson, "A theoritical analysis of recursive identification methods", Automatica, vol. 14, pp. 231-244, 1978.

[5] B. Friedlander, "A recursive maximum likelihood algorithm for ARMA spectral estimation", IEEE Trans. on Information Theory, vol. IT-28, no. 4, pp. 639-646, 1982.

[6] B. Friedlander, "The over-determined recursive instrumental variable method", IEEE Trans. on Automatic Contr., vol. AC-29, no. 4, pp. 353-356, 1984.

[7] N. J. Bershad, "Analysis of the normalized LMS algorithm with Gaussian inputs", IEEE Trans. on ASSP, vol. 34, no. 4, pp. 793-806, 1986.

[8] K. J. Astrom and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records", Proc. 2nd IFAC Symp on theory of self adpt. sys., pp. 96-111, 1965.

[9] B. Friedlander, "A modified prefilter for some recursive parameter estimation algorithms", IEEE Trans. on Automatic Control, vol. AC-27, no. 1, pp. 232-235, 1982.

[10] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes", IEEE Trans. on Automatic Control, vol. AC-22, no. 4, 1977.

[11] A. H. Gray and J. D. Markel, "Distance measures for speech processing", IEEE Trans. on ASSP, vol. 24, no. 5, pp. 380-391, 1976.