ON THE REPRESENTATION OF TIME–VARYING LPC PARAMETERS BY CUBIC SPLINES WITH VARIABLE KNOTS

M.J. Flaherty

Telecom Australia Research Laboratories

ABSTRACT – The modelling of log area coefficients by cubic splines with variable knots is discussed. Results are presented which compare variable and uniform knot modelling for a connected digit utterance spoken over a telephone handset using least squares error and spectral difference measures.

INTRODUCTION

Speech waveforms exhibit various degrees of non–stationarity and as a result a uniformly sampled, bandlimited approximation of model parameters is highly inefficient (Atal, 1983). Observations of model parameters such as reflection coefficients or log area ratios will show regions where parameters vary rapidly and other regions where parameters vary more slowly. The use of uniformly sampled parameters requires the sampling rate to be sufficiently high in order to cater for regions of more rapid change. One means of overcoming this problem is to use basis functions which adapt to the degree of local stationarity.

Polynomial splines belong to the class of piecewise polynomials with specific continuity constraints and have shown to be a useful basis for non–parametric (Flaherty, 1982) and parametric (Flaherty, 1988) representations. The rationale for selecting polynomial splines is outlined in (Flaherty, 1988). This paper presents results on the application of least squares cubic splines with variable knots to the problem of efficient speech model parameter representation. The resulting knot placement is equivalent to finding a set of basis functions which are adapted to match the degree of local stationarity. Conversely, the piecewise polynomial segments are of shorter duration in regions where parameters vary rapidly and of longer duration where the parameters vary slowly.

TIME–VARYING MODEL

A time–varying LPC model can be written as

$$y(n) - \sum_{k=1}^{p} a_k(n).y(n-k) = g(n).\varepsilon(n) \qquad (1)$$

where $a_k(n)$ are the time–varying lpc coefficients,
$g(n)$ is the time varying gain,
$\varepsilon(n)$ is the excitation.

electret and carbon microphones (conventional telephone handsets). The data was digitized with a 16 bit A/D converter at a sample rate of 8 kHz.

The log area ratio coefficients were obtained using autocorrelation LPC analysis with a 256 point sliding Hamming window. A slide interval of 32 samples (4 ms) was employed to obtain reflection coefficients which were used to obtain the log area ratio coefficients. A slide interval of 16 samples was initially used but did not appear to provide any advantage for the window size chosen. A larger slide interval reduced the computational load associated with the determination of the variable knot locations.

Analysis was carried out for uniform and variable knots with 10, 15 and 22 knots per utterance. A spectral distance measure used in this study is given by (Turner and Dickson, 1977) as:

$$\text{SPDIFF} = (10/\ln 10).\left[2. \sum_{k=1}^{p} (c_k - \hat{c}_k)^2 \right]^{1/2} \text{ dB} \qquad (3)$$

where c_k and $\hat{c}_k$ are the original and approximated cepstral coefficients.

This spectral distance measure is evaluated at the sample rate of the original lpc coefficients and presented as a function of time.

Normalized least squared errors were obtained for each log area ratio coefficient and defined by:

$$\text{normaized error} = \frac{\sum_{n=1}^{N} \left[\alpha_i(n) - S_i(\tau,n) \right]^2}{\sum_{n=1}^{N} \alpha_i(n)^2} \qquad (4)$$

DISCUSSION AND CONCLUSION

Figure 1a shows the variable knot approximation of the first and second log area ratio coefficient while figure 2a shows the same approximation with uniform knots. The variable knot approximation clearly provides a much better approximation in regions of rapid parameter change. This is particularly noticeable in the vicinity of sample 2500 corresponding to the onset of the word seven. In fact in the region of all onsets, the variable knot approximation shows significantly better results. This is achieved by the clustering of knots in the vicinity of rapid changes in log area ratio coefficients. The more stationary regions corresponding to voiced speech have a much wider knot spacing. Figures 1b and 2b show the spectral difference plots which quantify the improvement in spectral approximation.

Tables 1 and 2 show the normalized least squares approximation error for uniform and variable knot placement averaged over all the utterances as a function of number of knots and log area coefficient number. The results indicate that, from the point of view of least squares error, approximately twice the number of knots are required for uniform knot approximation to result in the same squared error. The results also indicate that the benefits of variable knot placement decreases with

It is common to approximate the time–varying parameters as being constant over an appropriate analysis window. However, by choosing appropriate basis functions to represent the parameters one can produce a set of linear equations and solve for the time–varying parameters. This problem has been solved using B–splines on uniform knots (Flaherty, 1988). Unfortunately the resulting linear system is large and the basis functions do not adapt to local signal stationarity. Splines can be used to adapt to local signal stationarity by allowing variable knot placement. The large system of linear equations then become a large system of highly non–linear equations which are not computationally tractable.

An alternative approach is to obtain an approximation of the time–varying parameters using conventional lpc analysis with a sliding window updating the parameters every few milliseconds. The parameters can then be modelled with variable knot splines resulting in a more computationally tractable problem. This is the approach adopted in the current work.

VARIABLE KNOT APPROXIMATION

It is not desirable to directly approximate the linear prediction coefficients for reasons of filter stability. A more suitable set of parameters for approximation are the log area ratio coefficients (Markel and Gray, 1976). The approximation problem dealt with here is to minimize the squared error given by:

$$\text{error} = \sum_{n=1}^{N} \left[\alpha_i(n) - S_i(\tau,n) \right]^2 \qquad 1 < i < p \tag{2}$$

where p is the predictor order,and $S_i(\tau,n)$ is a cubic spline with knot placement τ approximating the ith log area ratio coefficients $\alpha_i(n)$.

The cubic spline can be considered as a piecewise (cubic) polynomial with continuity of the approximant and its first and second derivative across the breakpoints, τ. In general a cubic polynomial is represented by four coefficients, however, the continuity constraints of the cubic spline result in K+2 coefficients defining a cubic spline over K knots for a given knot structure. The resulting cubic spline can be represented by K+2 B–spline basis functions which have compact support.

Given a fixed knot structure the problem is easily solved by forming the (linear) normal equations and solving a banded, positive definite system. Although there are potential gains to be obtained by allowing variable knots such as efficiency of representation, the problem becomes nonlinear. The algorithm adopted in the current work finds local minima and not global minima. The starting point is to assume a uniform knot spacing then optimize the position of each knot in turn using an established algorithm (de Boor and Rice, 1968). The solution is a brute force approach and requires the solution of a linear least squares problem at each stage of the iteration. After all the knot positions are individually optimized, the process is repeated until a termination criterion is met.

EXPERIMENTAL PROCEDURE

For the purposes of preliminary investigation, test data was obtained form three male and three female speakers speaking the connected digit utterance "6758". The data was recorded through a local PABX with telephone bandwith and a mixture of increasing log area coefficient number. These conclusions are supported by the spectral difference measure described earlier.

ACKNOWLEDGEMENT

The permission of the Executive General Manager, Telecom Australia Research Laboratories, to publish this paper is hereby acknowledged.

REFERENCES

Atal, B.S., "Efficient Coding of LPC Parameters By Temporal Decomposition", Proc. Int. Conf. on Acoustics Speech and Signal Processing, Boston, 1983, pp. 81–84.

de Boor, C. and Rice, J.R.,"Least Squares Cubic Spline Approximation, II – Variable Knots", CSD TR 21, Purdue University, West Lfayette, April, 1968.

Flaherty, M.J., "Time–Frequency Analysis of Speech Via Modulated Splines", Ph.D Thesis, Univ. of QLD., 1982.

Flaherty, M.J.,"Application of Polnomial Splines to a Time–varying Autoregressive Model of Speech", Proc. Int. Conf. on Acoustics Speech and Signal Processing. New York, 1988, pp. 2220–2223.

Turner, J. and Dickson, B.,"Linear Prediction Applied to Time–Varying All–Pole Signals", Proc. Int. Conf. on Acoustics Speech and Signal Processing, pp. 750–753.

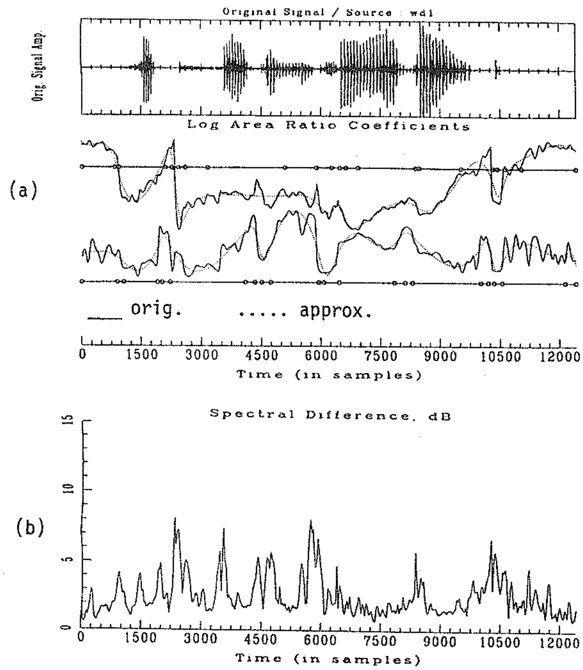

Figure 1. Approximation with variable knots.

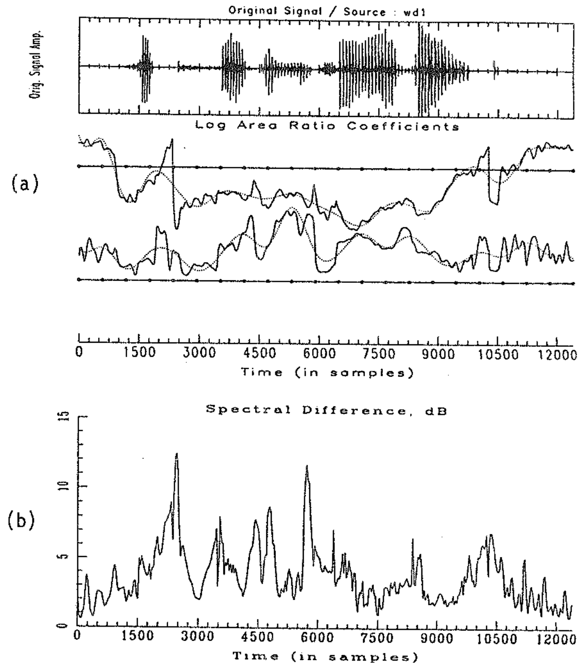

Figure 2. Approximation with uniform knots.

	Number of Knots		
Coeff	22	15	10
1	0.284199E-01	0.450116E-01	0.718623E-01
2	0.310227E-01	0.589218E-01	0.103571
3	0.498793E-01	0.772456E-01	0.113815
4	0.723164E-01	0.104293	0.148310
5	0.128277	0.191929	0.259726
6	0.111854	0.148775	0.195857
7	0.204824	0.273161	0.339570
8	0.999285E-01	0.130339	0.154567
9	0.180501	0.225883	0.291427
10	0.153079	0.192837	0.257589
11	0.247202	0.297198	0.370758
12	0.240643	0.294072	0.369120

TABLE 1. Average least squares error, Variable knots.

	Number of Knots		
Coeff	22	15	10
1	0.491369E-01	0.733979E-01	0.969564E-01
2	0.562538E-01	0.984204E-01	0.150315
3	0.831634E-01	0.104102	0.144492
4	0.108245	0.142770	0.187349
5	0.176426	0.253102	0.313922
6	0.146055	0.177600	0.221416
7	0.283039	0.317351	0.376827
8	0.126635	0.159358	0.173523
9	0.220825	0.293167	0.344025
10	0.188918	0.249169	0.292022
11	0.291707	0.356619	0.409121
12	0.304559	0.373348	0.411519

TABLE 2. Average least squares error, Uniform knots.