SUBBAND AUTOREGRESSIVE MODELLING FOR SPEECH ENHANCEMENT

Brady Laska[1], Rafik Goubran[1], and Miodrag Bolić[2]
[1]Dept. of Systems and Computer Engineering, Carleton University, Ottawa, Canada, {laska,goubran}@sce.carleton.ca
[2]University of Ottawa School of Information Technology and Engineering, Ottawa, Canada mbolic@site.uottawa.ca

1. INTRODUCTION

The speech waveform can be efficiently represented as an autoregressive (AR) process. Speech AR modelling is referred to as linear predictive coding (LPC) because the current speech signal sample $x(n)$ is represented as a linear combination of previous samples:

$$x(n) = \sum_{k=1}^{p} a(k)x(n-k) + u(n) = \mathbf{a}^T \mathbf{x}(n) + u(n), \quad (1)$$

where $\mathbf{a}^T = [a(1),\ldots,a(p)]$ is the vector of autoregressive model coefficients, $\mathbf{x}^T(n) = [x(n-1),\ldots,x(n-p)]$ is the signal vector, p is the model order and $u(n)$ is the excitation sequence. The use of AR models for speech signals has a physiological justification as $u(n)$ corresponds to the excitation from the lungs and the filter defined by the AR model coefficients corresponds to the all-pole vocal tract filter. The roots of the AR coefficient polynomial define the resonances of the filter which produce the characteristic formant peaks in the speech spectrum.

The parametric form of the AR model provides an efficient and low-variance representation of the speech signal spectrum. This allows for substantial compression gains in speech communications, and can also be applied to speech signal enhancement. Additive background noises - such as building ventilation system or in-car road noise - reduce speech intelligibility for human listeners and degrade the performance of automated speech and voice recognition systems. Speech enhancement algorithms attempt to remove the additive noise without distorting the desired speech signal. If the AR parameters of the clean speech signal are known, and the excitation and measurement signals are white Gaussian noise, Equation (1) can be arranged into a linear state-space form. This allows the Kalman filter equations to be used to obtain the minimum mean-square error (MMSE) optimal estimate of the clean speech waveform (Paliwal and Basu, 1987). By enforcing an AR model structure, Kalman filter speech enhancement provides high quality enhanced speech with natural sounding residual noise.

2. SUBBAND AR MODELLING

The rise of digital networks has enabled the emergence of systems transmitting wideband (16 kHz sampling rate) speech. While wideband speech improves the listener experience, it creates problems for AR speech modelling. The computational complexity of systems using AR speech models grows rapidly as the model order increases. Also, high-order models require more data to reliably estimate the parameters, but speech signals are only stationary over a short time. Low-order models are therefore desirable, however very high-order models are required to capture the pitch structure (Puder, 2006), and the deep troughs and steep slopes of the wideband speech spectrum.

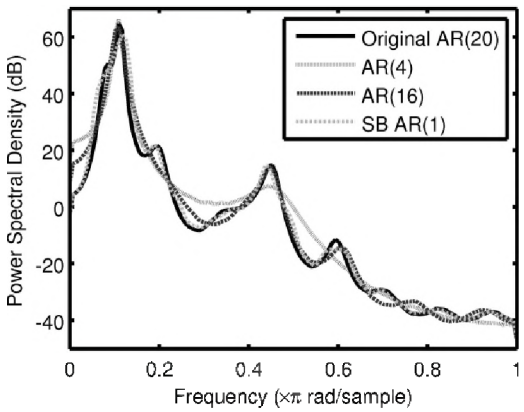

Fig. 1. Spectra obtained from fullband and subband AR modelling of an AR (20) process.

An alternative to using a single high-order AR model is to use a filterbank to decompose the speech signal, and to model each subband channel with a very low-order AR model. Since the model parameters and energy level of each band are determined independently, subband AR models need not exhibit the same spectral smoothness of fullband models, and may permit better modelling of steep spectral slopes and troughs. Furthermore, as the processing of each subband signal is carried out at a decimated (time-reduced) rate, the number of computations per unit time may be decreased. It has been demonstrated (Rao and Pearlman, 1996) that with ideal filterbanks, subband AR models can achieve lower modelling error. Here we investigate the performance of realizable filterbanks.

To compare the performance of low-order fullband and subband AR modelling of complex signals, an AR(20) process was generated by passing white noise through an all-pole filter measured from a segment of voiced speech.

Fullband AR(4) and AR(16) models were fit to the signals, and an AR(1) model was fit to each band of a 16-band cosine modulated filterbank, designed using the approach in (Lin and Vaidyanathan, 1998). The resulting spectral estimates are shown in Fig. 1. The AR(4) curve shows that when low-order fullband AR modeling is used to estimate a higher order process, the smooth fitting of a curve between the poles causes the signal energy between the poles to be significantly over-estimated. While the AR(16) model is an improvement over the AR(4), the subband estimate still provides the closest fit. The root mean squared error between the true estimated signal spectra for the AR(4), AR(16) and subband AR(1) models are 6.56 dB, 4.16 dB and 3.32 dB respectively.

3. SUBBAND KALMAN FILTERING

In the context of Kalman filter speech enhancement, the over-estimation of the spectral troughs by the low-order fullband AR models leads to increased residual noise, as the noise between spectral peaks is treated as speech. If the residual noise is sufficiently far from a formant peak, it will not be masked by the formant and will be perceptually noticeable. This problem is more prominent in wideband speech where the spectral dynamic range within a speech segment is higher, and there can be high and low frequency energy in the same frame. The better spectral modelling provided by a subband AR model may therefore be exploited in Kalman filter speech enhancement to achieve better noise reduction, and has been shown to offer good sound quality (Puder, 2006). As an added benefit, the subband signal decomposition enables frequency-dependent processing strategies. Speech formant peaks are concentrated in the low-frequency regions, while the high frequency regions are spectrally flatter; using different model orders in different bands could provide a further complexity reduction without degrading the model estimate.

Simulations were carried out to compare fullband and subband Kalman filter enhancement of a speech signal degraded by white Gaussian noise at an overall SNR of approximately 10 dB. Fullband AR(8) and AR(16) configurations were tested as were two subband configurations: one using AR(1) models in all bands, and another using AR(1) models in the lower 8 bands and AR(0) models in the upper 8. Table 1 presents the signal to noise ratio (SNR) results and Fig. 2 presents the spectra of a voiced speech segment obtained from the enhanced output. The improved modelling of the spectral troughs allows the subband configurations to achieve higher noise suppression, especially in the higher frequencies.

4. DISCUSSION

Wideband speech signal spectra can possess multiple peaks and deep troughs within the same short time segment. This diverse spectral character motivates the use of frequency-dependent processing strategies such as subband AR modelling. Kalman filter speech enhancement was used to demonstrate the potential benefits of this approach. In addition to offering better modelling performance at a lower complexity than the same order fullband model, subband systems allow the designer to vary the model parameters with frequency. A heterogeneous processing strategy using different model orders in high and low bands was shown to offer comparable noise reduction at a reduced complexity.

Table 1: Segmental and overall SNR scores.

Configuration	Segmental SNR	Overall SNR
Noisy Signal	0.96 dB	10.32 dB
Fullband AR(8)	5.21 dB	15.27 dB
Fullband AR(16)	5.28 dB	15.33 dB
Subband AR(1)	7.48 dB	18.25 dB
Subband AR(1/0)	7.46 dB	18.21 dB

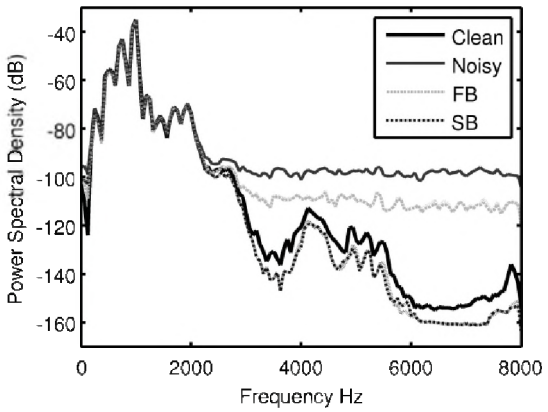

Fig. 2. Spectra of voiced speech segment from clean, noisy and fullband and subband Kalman filter enhanced signals.

REFERENCES

Lin, Y. and Vaidyanathan, P. (1998). A Kaiser window approach for the design of prototype filters of cosine modulated filter banks, IEEE Signal Proc. Letters, vol. 5. no. 6. pp. 132-134.

Paliwal, K., Basu, A. (1987). A speech enhancement method based on Kalman filtering. Proc. IEEE ICASSP, pp. 177-180.

Puder, H., (2006). Noise reduction with Kalman filters for hands-free car phones based on parametric spectral speech and noise estimates. In: Topics in Acoustic Echo and Noise Control, Springer.

Rao, S., Pearlman, W., (1996). Analysis of linear prediction, coding, and spectral estimation from subbands. IEEE Trans. Info. Theory 42 (4), pp. 1160-1178.

ACKNOWLEDGEMENTS

This work was funded in part by OGS, NSERC and Siemens AG.