VARIABLE BIT RATE ADPCM VIA ARITHMETIC CODING*'

*Craig R. Watkins**, *Sam Crisafulli, Robert R. Bitmead, Robert J. Orsi*

Department of Systems Engineering, Cooperative Research Centre for Robust and Adaptive Systems,
Australian National University, Canberra ACT 0200, Australia

ABSTRACT

We discuss the use of a variable bit rate ADPCM system for speech coding. We present an ADPCM system, using Arithmetic Coding and Kalman Filtering techniques, that can code signals at very low bit rates, where standard fixed bit rate ADPCM systems fail. The proposed system, at an average bit rate of 16 kbps, produces a reconstructed speech signal that by informal listening tests has been judged to be of significantly higher quality than the LD-CELP 16 kbps speech coding standard. We believe that the output quality is comparable to that of LD-CELP at an average rate of around 12 kbps, and the quality at 8 kbps is very promising. Other advantages of the scheme are a reduction in complexity over the LD-CELP standard, and an increase in design flexibility.

1. INTRODUCTION

Due to the success of CELP (Code Excited Linear Prediction) approaches for coding of speech and other signals, ADPCM (Adaptive Differential Pulse Code Modulation) has not received widespread attention in recent years. Indeed, for toll quality coding of speech, ADPCM would appear to be at its limit with the CCITT G.721 32 kbps standard. Below four bits per sample, traditional fixed rate ADPCM systems quickly encounter performance versus stability problems which many CELP approaches are able to avoid.

We examine closely the performance of a variable bit rate ADPCM system, and attempt to make a comparison of its performance and complexity with CELP approaches. The use of the variable bit rate overcomes the performance versus stability problems that are observed within the fixed rate ADPCM systems, and allows very low average bit rates to be obtained without stability problems.

We discuss the stability problem of ADPCM in the next section, leading to the introduction of our proposed variable bit rate system. After some coverage of the performance of the system with a standard linear predictor, we introduce a Kalman predictor, and show an approach to reduce the computational complexity. We discuss the use of perceptual weighting within the system, and bit error resynchronization for the Arithmetic Coding (AC) data stream. We also mention briefly some concerns related to possible applications of the scheme.

*Work supported by the Australian Telecommunications and Electronics Research Board (ATERB) and the International Business Unit of Telstra Corporation LTD. **The authors wish to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Australian Commonwealth Government under the Cooperative Research Centres Program.

2. ADPCM DESTABILIZATION EFFECTS

Fixed rate ADPCM systems such as that shown in block form in Figures 1 and 2, generally consist of both adaptive quantizers and adaptive predictors. For speech coding, this adaptation is used to account for the wide variations in the input signal statistics.

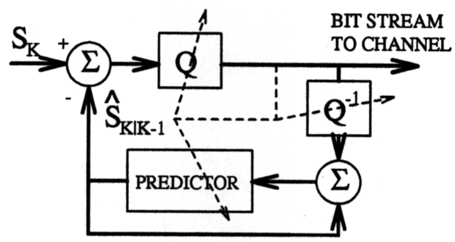

Figure 1. Fixed Rate ADPCM Encoder Block Diagram

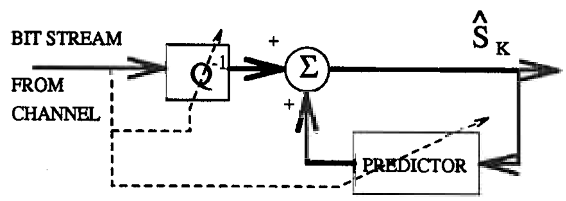

Figure 2. Fixed Rate ADPCM Decoder Block Diagram

Consideration of the ADPCM encoder reveals a feedback loop, and it is easy to see that depending on the adaptation strategies for the quantizer and predictor, stability problems may exist. Several authors have noted the existence of stability problems [6, 7, 8], and some studies of stability have been undertaken [9, 10, 11].

Going to a 24 kbps ADPCM system shows a significant decrease in performance over the 32 kbps system. A 16 kbps ADPCM system gives extremely poor performance, with evidence of severe stability problems being observed in the output. Jayant and Noll[1] show it is possible to tune the adaptive quantizer to effectively eliminate this stability problem, but the cost is that the performance of the 16 kbps ADPCM system is extremely poor. Tabulations of 'loading factors' are provided by Jayant and Noll, to be used with adaptive quantizers of various numbers of bits to give a reasonable stability/performance compromise.

Previous work by the authors[11], has shown how the stability of the ADPCM loop is closely linked to the rate of quantizer step size decrease. Slower adaptation in the quantizer step size is shown to improve stability, but with standard ADPCM systems, this usually comes at a performance cost. The next section deals with an ADPCM system that has improved stability due to elimination of the quantizer step size adaptation, but that is able to maintain performance through the use of a variable bit rate.

3. VARIABLE BIT RATE ADPCM

The stability problems with the fixed rate ADPCM systems are related to the fact that the fixed rate quantizer has conflicting requirements. It must have levels spaced closely together to obtain good coding performance, but must have levels spaced far apart to ensure stability by being able to track rapid changes in input signal statistics. For lower bit rate systems, these two requirements become increasingly difficult to meet simultaneously.

Lifting the restriction of a fixed rate system is one approach that can help to relieve the stability problems. Moving away from a sample-by-sample ADPCM approach to a vector basis CELP approach is another. CELP approaches have been well studied and extremely successful in practice, but codebook searches are required and these imply a significant computation cost. We investigate the variable bit rate ADPCM approach.

The system we have constructed for simulation is shown in block form in Figure 3. Here it can be seen that the basic feedback loop of ADPCM is intact. However, the quantizer within the loop now has an many levels (practically infinite) to avoid overload distortion, and Arithmetic Coding is used as post-processing of the quantizer output data stream. We are thus dealing with an instantaneous variable bit rate system. This system we refer to as Arithmetic Coding ADPCM (AC-ADPCM). It is important to note that the "infinite" number of levels in the quantizer certainly does not imply zero granular distortion, only that the quantizer does not have any overload regions.

Arithmetic Coding[2] is a practically optimal entropy coding scheme, that is used here to encode the quantizer output with only the number of bits required by information theoretic considerations, based on the probability of that quantization level being used. The prime difference with Arithmetic Coding to Huffman coding is that it does not suffer from the disadvantage of requiring each source

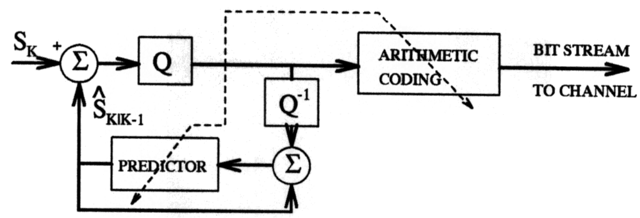

Figure 3. Arithmetic Coding ADPCM Encoder Block Diagram

symbol to be encoded with an integral number of bits.

The system shown in Figure 3 is similar in some ways to one investigated by Howard and Vitter[4, 5] for DPCM compression of images. However, there are some significant differences for an ADPCM system for coding speech.

A quantizer with no overload distortion means that we have ample ability to track rapid changes in input signal statistics, without feeding larger errors into the predictor, and causing instability as with the Jayant adaptive quantizer in fixed rate ADPCM.

As well as obtaining an advantage over fixed rate ADPCM systems through the elimination of the stability problems, the system also has the advantage of exploiting inactive periods of input to reduce the average bit rate. Of course, it is dependant upon the application as to whether this type of saving is possible or desirable.

The linear predictor shown in Figure 3 is updated in a backwards adaptive manner, using autocorrelation analysis, and Levinson-Durbin recursion, similar to what is done in LD-CELP (CCITT G.728). The arithmetic coding block uses source symbol probabilities based on the assumption that the input to the quantizer can be modelled as a Laplacian distribution, with a backwards adaptive variance estimate.

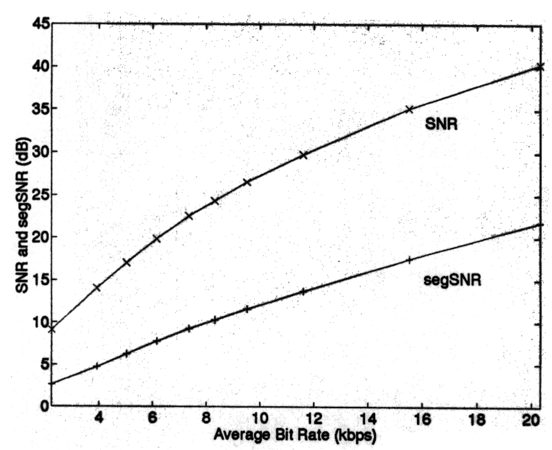

Figure 4. SNR and segSNR versus Average Bit Rate

Simulations and informal listening tests show that the performance at an average bit rate of 16 kbps is quite good, certainly a fantastic improvement over 16 kbps fixed rate ADPCM. A plot of the SNR and segmental SNR versus the average bit rate is shown in Figure 4. As the average bit rate is reduced, the performance degrades, but appears to do so in a graceful manner. Unfortunately, even at 16 kbps the subjective quality does not seem to be toll quality, as the quantization noise is clearly audible, especially in the higher frequencies. The next subsection discusses the use of Kalman filtering techniques in an attempt to reduce this audible noise.

3.1. Kalman Prediction

A Kalman predictor is able to obtain an improvement over the standard linear predictor by exploiting knowledge of the quantization noise statistics, when the quantization noise is modelled by zero mean white noise. This is ignored in the standard linear predictor, leading to non-optimal linear prediction. The Kalman filter provides 'optimal' linear predictor performance, but at a high computational cost. We exploit the use of smoothing within the Kalman predictor to reduce the computational cost to a more manageable level, without sacrificing performance significantly.

The all-pole signal model we use for speech is

$$S_k = a_1 S_{k-1} + a_2 S_{k-2} + a_3 S_{k-3} + \cdots + a_N S_{k-N} + w_k, \quad (1)$$

where S_k are the speech samples, N is the model order, a_i; $i = 1, \ldots, N$ are the predictor coefficients and w_k is the excitation sequence. Based on this signal model and information available up to sampling instant $k-1$, the best linear prediction of the sample at time k is given by

$$\hat{S}_{k|k-1} = a_1 S_{k-1} + a_2 S_{k-2} + a_3 S_{k-3} + \cdots + a_N S_{k-N}. \quad (2)$$

Of course, for signal coding purposes, the samples $S_{k-1} \cdots S_{k-N}$ are not available at the decoder, so standard linear predictors use instead a prediction of the form

$$\begin{aligned} \hat{S}_{k|k-1} &= a_1 \hat{S}_{k-1|k-1} + a_2 \hat{S}_{k-2|k-2} + a_3 \hat{S}_{k-3|k-3} \\ &\quad + \cdots + a_N \hat{S}_{k-N|k-N}, \end{aligned} \quad (3)$$

where $\hat{S}_{k-1|k-1} \cdots \hat{S}_{k-N|k-N}$ are the reconstructed (filtered) sample values that are available at the decoder.

The Kalman predictor, as presented in [3], uses smoothed estimates of the past signal samples, to form a prediction of the form

$$\begin{aligned} \hat{S}_{k|k-1} &= a_1 \hat{S}_{k-1|k-1} + a_2 \hat{S}_{k-2|k-1} + a_3 \hat{S}_{k-3|k-1} \\ &\quad + \cdots + a_N \hat{S}_{k-N|k-1}. \end{aligned} \quad (4)$$

Unfortunately, Kalman filtering of this sort requires significant computation resources. However, from [12], we are able to exploit the fact that most of the smoothing gain is achieved from only the first few lags, to obtain good performance with only moderate computation requirements. The prediction that we actually use is of the form

$$\begin{aligned} \hat{S}_{k|k-1} &= a_1 \hat{S}_{k-1|k-1} + a_2 \hat{S}_{k-2|k-1} + \cdots + a_n \hat{S}_{k-n|k-1} \\ &\quad + a_{n+1} \hat{S}_{k-n-1|k-2} + a_{n+2} \hat{S}_{k-n-2|k-3} + \cdots \\ &\quad + a_N \hat{S}_{k-N|k-N+n-1}, \end{aligned} \quad (5)$$

where n is the number of lags to which smoothing is performed. It is only by being able to reduce the complexity in this way that the Kalman filtering approach becomes practical.

In terms of SNR and segmental SNR, the gain obtained through the use of the Kalman predictor is fairly minor, at around 1.5 dB SNR, and 0.8 dB in segmental SNR. However, the subjective improvement obtained is quite significant, with almost all the audible quantization noise being removed at an average bit rate of 16 kbps.

3.2. Perceptual Weighting

Perceptual Weighting is a vital component of many speech coding systems, and is found to give significant subjective performance improvements in the variable bit rate ADPCM system above. We use an AR filter for perceptual weighting, and thus the inverse perceptual weighting filter is simply a MA structure.

The coefficients of the perceptual weighting filter are bandwidth expanded versions of the coefficients of a low order linear predictor. This is somewhat similar to LD-CELP, where the coefficients for the perceptual weighting are obtained from partly completed Levinson-Durbin recursion used to obtain the predictor coefficients.

4. OTHER ISSUES

4.1. Design Flexibility and Computation

We have mentioned above how the output from the AC-ADPCM system compares to that from the LD-CELP CCITT G.728 standard. Another very important practical concern is the issue of computational complexity. The AC-ADPCM approach avoids the codebook search required with CELP systems, which can represent a significant computation saving. However, it would seem that there is no significant difference between the two schemes as far as the other elements of the system are concerned, such as the backwards adaptation of the predictors.

This is not necessarily true, as the AC-ADPCM system has design flexibility in trading off computatal complexity against output quality and average bit rate. CELP approaches are often able to trade some complexity for output quality, but it is quite difficult to have any significant flexibility in the output bit rate.

The variable bit rate approach also has another possible advantage. It may be useful to have a quality control 'knob', where the quality can be increased on demand, of course at a cost of a higher average bit rate. Other applications may utilize the quality 'knob' for the system to demand a lower bit rate, and hence lower output quality. The whole issue of bit rate control stems from these considerations, but we do not investigate them in any depth here.

4.2. Applications

Applications for such a coding system, with an instantaneous variable bit rate, are mainly related to storage systems, and where a variable rate transmission system is available. For storage, usually it is safe to assume low bit error probability, so the arithmetic coding approach can be practically useful, while only needing a bare minimum of overhead bits for resynchronization.

ATM (Asynchronous Transfer Mode) transmission of speech is an area for which many variable bit rate systems have been touted. However, as the basic aim of ATM networks is to provide broadband communications, it is unclear as to whether the compression of speech is likely to be of significant consequence. Overall delay requirements also make it unclear as to whether the ATM cells can be filled with a single data stream.

Acknowledging the uncertainty with respect to the ATM application, we note that the AC-ADPCM approach would appear to be suitable for ATM use. The problem of packet loss within an ATM system means that we are likely to need significantly higher resynchronization overhead than for the storage application, but this does not necessarily imply a huge additional cost. The AC-ADPCM scheme also allows for the use of an embedded coding approach, where some packets are designated as high priority packets, and receive the coarse quantizer information, with the fine detail going into the lower priority packets.

Variable bit rate mobile communications is another area of interest for the scheme, although currently it is unclear as to how to overcome the transmission error problems that exist with AC. As wireless channels for mobile communications have high variations in transmission conditions, this is a significant problem.

5. CONCLUSIONS

The use of variable rate ADPCM via Arithmetic Coding has been shown to practically eliminate the stability problems apparent with fixed bit rate ADPCM systems at lower bit rates than the 32 kbps G.721 standard.

The variable bit rate AC-ADPCM system is also able to exploit inactive periods during speech utterances, to give a relatively low average bit rate. Although quite reasonable performance is achieved at an average rate of 16 kbps, the system with a standard linear predictor provides an output that is not toll quality, due the the presence of significant levels of audible quantization noise.

The use of Kalman prediction practically eliminates the audible quantization noise present at an average bit rate of 16 kbps, and results in high speech quality at this rate. At an average rate of about 12 kbps, the output speech quality is comparable to that of LD-CELP at 16 kbps, and at an average rate of 8 kbps, the speech quality is slightly below LD-CELP, with noticeable differences occurring between the voiced and unvoiced sections. Further work promises to improve performance at the 8 kbps average rate to near toll quality.

Immediate applications for the scheme are in speech storage. Here the variable bit rate, and the AC bit error problems do not constitute a disadvantage. Other potential applications include speech compression for ATM networks, and variable rate mobile telephony.

REFERENCES

[1] N. S. Jayant, and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice-Hall, New Jersey, 1984.

[2] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic Coding for Data Compression", Communications of the ACM, Vol. 30, No. 6, June 1987.

[3] S. Crisafulli, J. D. Mills, and R. R. Bitmead, "Kalman Filtering Techniques in Speech Coding", Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, San Francisco, March 1992.

[4] P. G. Howard, and J. S. Vitter, "New Methods for Lossless Image Compression Using Arithmetic Coding", Proceedings of the 1991 IEEE Data Compression Conference, Snowbird, Utah.

[5] P. G. Howard, and J. S. Vitter, "Parallel Lossless Image Compression Using Huffman and Arithmetic Coding", Proceedings of the 1992 IEEE Data Compression Conference, Snowbird, Utah.

[6] M. Honda, and F. Itakura, "Bit Allocation in Time and Frequency Domains for Predictive Coding of Speech", IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-32, No. 3, June 1984.

[7] V. Iyengar, and P. Kabal, "A Low Delay 16 kb/s Speech Coder", IEEE Trans. Signal Processing, Vol. 39, No. 5, May 1991.

[8] S. Crisafulli, G. J. Rey, C. R. Johnson, and R. A. Kennedy, "A Coupled Approach to ADPCM Adaptation", to appear IEEE Trans. Speech and Audio Proc., January 1994.

[9] O. Macchi, and C. Uhl, "Stability of the DPCM Transmission System", IEEE Trans. Circuits and Systems, Vol. 39, No. 10, October 1992.

[10] R. A. Kennedy, and C. R. Johnson Jr., "Encoder Stability Study for the 32 kbit/s CCITT G.721 ADPCM Standard based on a Simplified Error Model", Proceedings of The Second International Symposium on Signal Processing and its Applications, Gold Coast, Australia, August 1990.

[11] C. R. Watkins, R. R. Bitmead, and S. Crisafulli, "Destabilization Effects of Adaptive Quantization in ADPCM", Submitted to IEEE Trans. Speech and Audio Proc., December 1992.

[12] C. R. Watkins, S. Crisafulli, and R. R. Bitmead, "Reduced Complexity Kalman Filtering for Signal Coding", Presented at IEEE Intelligent Signal Processing and Communication Systems International Workshop, Sendai, Japan, 1993.