ISCA Archive
http://www.isca-speech.org/archive

4th European Conference on
Speech Communication and Technology
EUROSPEECH '95
Madrid, Spain, September 18-21, 1995

Low and Variable Bit-Rate Speech Coding for ATM Networks

S. D. Watson, B. M. G. Cheetham*, W. T. K. Wong# and A. V. Lewis#*
**Department of Electrical Engineering and Electronics, The University of Liverpool, LIVERPOOL. L69 3BX. U.K.*
#B. T. Laboratories, Martlesham Heath, IPSWICH. IP5 7RI. U.K

10.21437/Eurospeech.1995-166

ABSTRACT

This paper investigates the use of a variable low bit-rate speech transmission method for Asynchronous Transfer Mode (ATM) networks. A modified version of an existing 8kbit/s ACELP speech coder, making it an on-off source controlled variable bit-rate coder, referred to as AV-ACELP, has been adopted for this investigation. A possible adaptation of this coder to ATM is proposed and the effects of cell loss, cell delay and congestion control are examined. This paper discusses a suitable packetization method, ATM adaptation layer function and cell loss recovery scheme. Results obtained from this study are compared with results obtained from a standard 32kbit/s embedded ADPCM coder.

1. INTRODUCTION

Asynchronous Transfer Mode (ATM) is destined to be the predominant data transmission technique for networks of the future because of the growing demand for Broadband-ISDN services [2]. ATM is attractive for international markets and private PABX's owing to the versatility of the network which can be adapted to new services, and the ease of interworking with the future ATM public national networks. The support of voice traffic is essential for these future systems to ensure the success of this new network technology.

The justification for researching variable low bit-rate speech coding for ATM is cost efficiency. ATM requires specially adapted speech coders in order to exploit the statistical multiplexing capabilities of the network. Existing low bit-rate coding techniques can be adapted for this application by taking advantage of the on-off nature of telephone speech. Many such variable bit-rate channels can be statistically multiplexed onto one link for high speed digital transmission, yielding good bandwidth efficiency. However, ATM must still be able to send voice traffic with toll quality.

Two low bit-rate speech coders which may be suitable for adaptation to ATM are ACELP [3] and CS-CELP [4]. Both coders provide near toll quality coding performance and were under study by the ITU (formerly CCITT) for the 8kbit/s standard. The ACELP codec is chosen for this investigation, and has been modified to operate as a variable bit-rate (on-off) source controlled coder. The coder has been used to investigate the issues involved with integrating variable low bit-rate codecs into ATM networks. To enable the statistical multiplexing of ACELP coded voice channels, the GSM Voice Activity Detection (VAD) algorithm [5] was used to determine the active portions of the speech to be coded by ACELP. This resulted in a technique referred to as Active Voice ACELP (AV-ACELP). The use of the VAD allowed inactive speech frames to be detected and removed from the multiplexer. The quality of the ACELP coding technique was thus retained for each of the channels, while achieving higher over-all bandwidth efficiency. Further bandwidth savings can be made by using variable bit-rate coding for the active speech segments. However, the scope of this paper is limited to simple dual classification of the speech, i.e. classification into either active speech or silence.

The network model under study provides the ATM equivalent of Digital Circuit Multiplication Equipment (DCME)-Packet Circuit Multiplication Equipment (PCME) [6]. Here several input trunk channels are compressed by variable low bit-rate speech coders and packed onto a single bearer channel. This model assumes the use of the European standard high speed links of 2Mbit/s (E1) and 34Mbit/s (E3).

The issue of cell loss, which can occur with any ATM traffic, is of considerable importance when the quality of coded speech is being investigated. To reduce congestion during busy periods, a cell discarding scheme is used and a cell loss recovery technique is therefore required to reduce the perceptual distortion caused by this. The way in which low bit-rate speech coders react to typical cell losses may determine their suitability for ATM.

Finally, the traffic issues of cell delay, due to cell queues at the ATM nodes, are addressed by modelling the queues as M/D/1/K statistical models, described in [1]. This model is accurate for the network scenario of compressed voice multiplexing under study. Simulation results, obtained from AV-ACELP, are compared with results obtained for ADPCM [1]. The results from the traffic modelling lead to an easy way of assessing the viability of AV-ACELP for ATM transmission, as well as providing a comparison with coders that have already been implemented over ATM.

2. BI-MODAL SPEECH CODER

The two modes of operation are: silence, where no information is required, and active, where the full 8kbit/s ACELP coding is used.

It is generally understood that conversational telephone speech contains approximately 60% of silence and 40% of active speech bursts. If only the active speech bursts are coded, the channel capacity can be more than doubled. Toll quality speech transmission using this feature requires a Voice Activity Detector (VAD) which is robust to different types of background noise. Comfort noise must be injected to simulate the background noise effect at the decoder.

The Group Spécial Mobile (GSM) Voice Activity Detector [5] provides a reliable algorithm to distinguish between the active and silent segments of the input speech. Using this algorithm in conjunction with the ACELP coder produces a near toll quality variable bit-rate speech coder with a low average bit-rate.

The full functionality of ACELP is described in [3]. The coder is designed around the Code Excited Linear Predictive (CELP) coding model with coded speech sent every 10ms. The coding delay is much lower than many other coders at the same rate, thus giving it more flexibility for system designers. The special feature of ACELP is an algebraic innovation codebook structure which has many benefits in terms of storage, search complexity and search speed.

The GSM VAD [5] pre-processes the speech to determine whether or not the speech frame is active, and uses this decision to turn the encoder on or off. The VAD uses a 20ms speech frame to make its decision; this corresponds to two ACELP frames. The principles of the AV-ACELP coder are shown in figure 1.

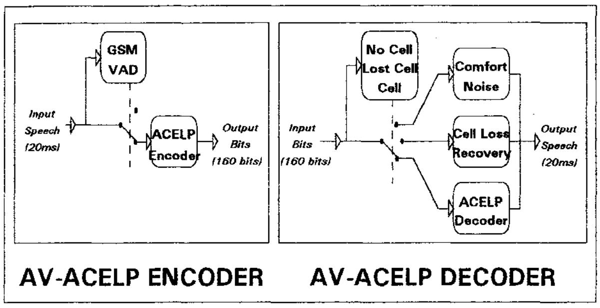

Figure 1 : AV-ACELP Encoder and Decoder Principles

AV-ACELP produces comfort noise at the decoder when it is notified that the encoder is inactive. This is done via the ATM network layers. Background noise is considered to be relatively stationary with power variations exhibiting small variance about a mean value. The decoder uses the linear predictive coefficients from the last decoded frame to spectrally shape the noise that is generated by exciting the synthesis filter by a random Gaussian vector. The output is scaled to the same power as that of the last decoded frame. Since the GSM VAD employs a technique of overhang to ensure that the speech frames are truly the onset of silence (not inter speech burst gaps), then it can be assumed that the power of the final frame is close to that of the mean power for the forthcoming silence period. This technique removes the discontinuities in power that would otherwise occur at the decoder and perceptually smoothes the transitions and silence periods.

3. ATM ISSUES

The ATM cell is 53 octets (bytes) long, with 48 octets of user information (*payload*) and 5 octets of control information (*header*). Two major drawbacks exist for voice transmission over ATM. One is cell delay and the second is cell loss, which can occur due to queue overflow (caused by congestion) or excessive variable cell delay.

ATM Adaptation Layers (AAL's) adapt the service information to the ATM stream. In essence, the AAL takes service information bitstreams from higher network layers and adapts them in such a way that the lower layer, the ATM layer, always sees the same structure. This is achieved by using some of the payload data bytes for the AAL, which contains information that is application specific for end-to-end communications. Thus different AAL protocols allow many varied traffic types to be transported through the network in an optimum way.

Variable bit-rate traffic requires certain parameters for its adaptation layer. Figure 2 shows a possible AAL for AV-ACELP, adopted for this investigation.

AV-ACELP Cell Construction

Header	AAL	Service Data Units (SDU)
5Bytes	8Bytes	40Bytes

AAL Construction

Seq. No. + Protection	Code Type	Noise Field	I T	Length Indicator	FEC	RESERVED
1Byte	4b	4b	2b	6b	1 Byte	4 Bytes

Figure 2 : ATM Adaptation Layer for the AV-ACELP.

A sequence number field will allow for the recovery of lost, misrouted or discarded cells. The coding type field will allow other low bit-rate coders to use this AAL. A noise field indicator provides a flag to show if the cell is the end of talkspurt, thus allowing the decoder to start comfort noise injection. An information type field will be able to indicate Beginning of Voice (BOV), Continuation of Voice (COV) and End of Voice (EOV). Since the traffic is variable, cell payloads may not be completely filled every time, and thus a length indicator giving the number of useful bytes in a partially filled cell is also a necessary field. Finally, the AAL header should be protected by a FEC field capable correcting minor errors. Four bytes are reserved for other possible purposes. The payload therefore has 40 bytes left for the four ACELP frames.

The total delay incurred by an ATM cell, excluding transmission delay, is equal to the input buffer delay plus coding and decoding delays. The buffering delay is the most prominent and is 45ms for an AV-ACELP cell. The transmission delay provides a large contribution to the overall delay. ATM cells experience variable delays while passing through queues at the switching nodes, and the decoder uses a build out delay buffer to remove the end-to-end jitter. A total end-to-end delay of up to 150ms one-way is considered reasonably acceptable for real-time voice applications [7] and is achievable with AV-ACELP over ATM.

The traffic characteristics of aggregate compressed voice cells statistically multiplexed exhibit burstiness properties that become less noticeable for higher speed links (more input sources allowed) and lower link loading. An M/D/1/K model, similar to the one used in [1], has been used to model the queuing characteristics of the multiplexers. No priority was assigned to any of the cells. When the queue exceeded the threshold, during times of congestion, then a cell was discarded. During the service intervals of the cells, a random Poisson number of cells, dictated by the length of these service intervals, arrived at the input of the queue. Thus the queue fluctuated on a service interval basis. This pseudo real-time model was simulated and validated against the M/D/1/K model by performing several thousand iterations. The pseudo real-time model was used to generate cell loss in the AV-ACELP and the results are given in section 5.

Mean cell delay for the E1 high speed link is modelled by the M/D/1/K model. These results are also given in section 5.

4. CELL LOSS RECOVERY METHOD

The work by Sriram *et. al.* [1] discussed the use of embedded coding to allow for graceful degradation of quality under cell loss conditions. At present AV-ACELP does not have this embedded property, and relies entirely on the extrapolation of parameters from previous cells.

Since each cell contains four AV-ACELP frames, simply repeating the last frame's parameters would impose excessive stationarity on the decoded speech. Therefore, successive frames are subtly altered with a progressively larger degree of bandwidth expansion of the linear predictive coefficients, gain tapering with a 3dB reduction per frame, and a randomised selection of innovation codebook excitation vector (pitch values remain the same). When the following frame is correctly received at the decoder, it is altered to smooth any transitions. This means that when a cell is lost, the recovery method estimates the lost cell while gradually reducing its audible perception.

AV-ACELP uses a cell loss regeneration technique that can deal with lost cells no matter how they occur. Conversely, the embedded ADPCM method has no recovery scheme for loss of high priority cells, which may occur due to unrecoverable header errors or excessive variable cell delay.

5. RESULTS

The M/D/1/K statistical queuing model was used to compare the 32kbit/s embedded ADPCM and the 8kbit/s AV-ACELP coders when applied to an ATM multiplexer. The same queue dimensions as used for the 1.536Mbit/s link in [1] were used for this simple comparison. The model was run for both E1 and the E3 link speeds.

Figure 3 shows the variation of mean cell loss against the number of input channels for both AV-ACELP and ADPCM over the E1 link. For mean cell losses below 1E-5 the fluctuations are probably due to the finite precision used by the simulation.

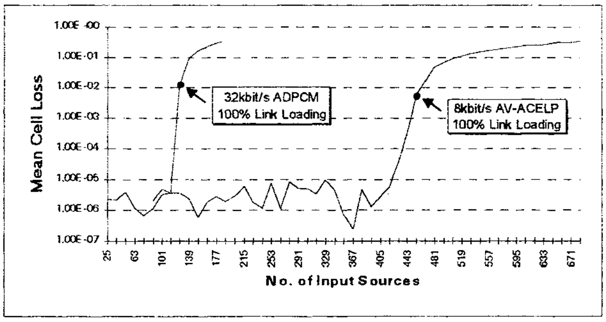

Figure 3 : Cell Loss vs. Input Sources for E1

As indicated in figure 3, for 100% loading of the output bearer channel (E1), the maximum number of sources supported by ADPCM with statistical multiplexing is 127, while AV-ACELP can support 460 input channels for the same conditions. Thus the additional compression gained by AV-ACELP allows 3.6 times the volume of traffic, possible from ADPCM, to pass on the E1 link. For these conditions both coders suffer approximately 1% cell loss, but recover well. At greater cell loss the embedded structure of the ADPCM significantly outperforms the AV-ACELP's cell loss regeneration scheme.

For E3 the trends are the same as above but the number of input sources is far greater.

The mean cell delay incurred by both codecs at a single ATM node against the number of input sources, for the E1 link, is shown in figure 4.

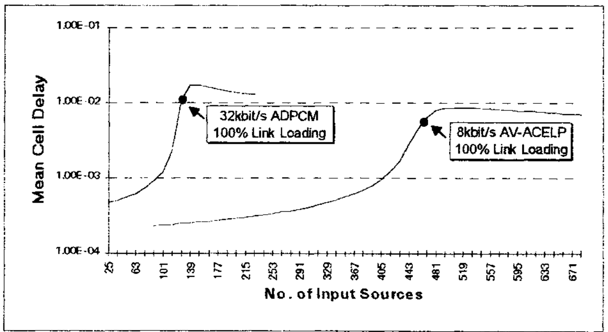

Figure 4 : Mean Cell Delay vs. Number of Sources for E1.

Figure 4 shows that the maximum mean cell delay, for AV-ACELP, is approximately 10ms. Assuming that the worst case scenario does not exceed twice this mean value, then the build out delay will be in the order of 20ms. This would mean that the number of cells lost due to excessive variable delay would become negligible with respect to cells loss as a result of congestion. When there are 138 input sources, on average an ADPCM cell experiences 17ms delay, while an AV-ACELP cell only experiences 250μs delay.

At E3 link speeds, the delay incurred by the two schemes have the same trends as shown in figure 4, but both have lower delays throughout reducing the delay observed by the user.

The AV-ACELP and the embedded ADPCM coders were both subjected to various cell loss scenarios and the decoded speech was recorded. These speech files were used for informal listening tests. Table 1 shows the percentage of listeners who thought the decoded speech, for the AV-ACELP, was either unintelligible or unacceptable for each of the seven different values of percentage cell loss. Table 2 shows the preferences of listeners for one coder or another under ten different cell loss conditions.

Cell Loss(%)	5	10	15	20	25	30	35
Unintelligible(%)	0	0	0	8	25	33	100
Unacceptable(%)	0	8	33	58	100	100	100

Table 1 : Intelligibility and Quality of AV-ACELP.

Cell Loss(%)	1	2	3	4	5	6	7	8	9	10
AV-ACELP(%)	8	25	8	16	42	25	16	25	0	0
ADPCM(%)	92	75	92	84	58	75	84	75	100	100

Table 2 : Quality Preference for AV-ACELP and ADPCM.

Table 1 indicates that AV-ACELP produces speech that most people consider to be unintelligible at cell losses of about 30% (142% link loading or 656 channels). In comparison ADPCM is known to produce intelligible speech up to, and a little beyond, 50% cell loss (200% link loading or 253 channels).

At 10% cell loss, the distortion produced by AV-ACELP renders the decoded speech unacceptable for a number of people, whereas ADPCM is considerably more acceptable at this cell loss.

AV-ACELP can provide good quality speech when the number of input channels forces heavy cell loss for ADPCM or renders it unusable (figure 3). However, table 2 demonstrates that although the majority of people preferred ADPCM to AV-ACELP with the same percentage cell losses, up to about 8%, there were people who preferred AV-ACELP and therefore the differences are not very marked. However, with cell losses greater than 8% there was a clear preference for ADPCM. It is important to realise that the nature of the distortion produced by the two codecs, under cell loss, is very different which makes the informal listening tests very subjective.

Further informal listening tests were performed that suggested that the same quality of speech, as ADPCM with 33% cell loss (150% link loading or 190 channels), can be achieved by AV-ACELP with between 10%-15% cell loss (113%-120% link loading or 520-552 channels).

6. CONCLUSIONS

For ATM speech coders, as the bit-rate is reduced more speech information is compressed into a single cell. Therefore, when this cell is lost, by whatever mechanism, an increasingly large amount of speech data has to be recovered by the cell loss recovery scheme. This makes cell regeneration a far more difficult task, and as a consequence, the quality achieved degrades dramatically with increasing link overload. However, schemes such as 32kbit/s ADPCM can produce a graceful quality degradation up to high link overload. Hence, if high channel capacity is required by using low bit-rate coders, then the degree of allowable overload is far less than would be acceptable for higher bit-rate coding schemes. However, improved cell loss recovery mechanisms will allow low bit-rate coders to adequately cope with higher degrees of cell loss due to congestion, while still maintaining acceptable quality.

The conclusions drawn from this paper give an insight into the problems faced by engineers developing low and variable bit-rate speech codecs for ATM.

ACKNOWLEDGEMENT

The authors acknowledge the support of BT Laboratories and EPSRC in this work. Thanks must also be given to Dr. K. Evans for his patience and comments.

REFERENCES

[1] K. Sriram, S. McKinney, and M. H. Sherif, "Voice Compression and Packetization in Broadband ATM networks", IEEE Journal on Sel. Areas of Comms., SAC-Vol. 9, No. 3, April 1991.

[2] M. De Prycker, "Asynchronous Transfer Mode - Solutions for Broadband ISDN", Second Edition, Ellis Horwood Series in Computer Communications and Networking, 1993.

[3] R. Salami, C. Laflamme, and J-P. Adoul, "ACELP Speech Coding at 8kbit/s with a 10ms frame : A Candidate for CCITT Standardization", IEEE Speech Coding Workshop, Canada, October 1993.

[4] A. Kataoka, T. Moriya, and S. Hayashi, "Conjugate Structure CELP Coder for CCITT 8-kbit/s Standardization candidate", IEEE Speech Coding Workshop, Canada, October 1993.

[5] D. K. Freeman, G. Cosler, C. B. Southcott and I. Boyd, "The Voice Activity Detector for the Pan-European Digital Cellular Mobile Telephone Service", ICASSP '89, S7.6, pp.369-372, May 1989.

[6] CCITT Recommendations - DCME (G.763) and PCME (G.765).

[7] J. G. Gruber, "Delay Related Issues in Integrated Voice and Data Networks", IEEE Trans. Commun., Vol. COM-29, pp. 787-800, June 1981.