AN ADAPTIVE DISCRETE COSINE TRANSFORMED VIDEOPHONE COMMUNICATOR FOR MOBILE APPLICATIONS

J. Streit and L. Hanzo

Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO17 1BJ, UK.

ABSTRACT

A highly bandwidth efficient, fixed but arbitrarily programmable rate, perceptually weighted Discrete Cosine Transform (DCT) based video communicator for quarter common intermediate format (QCIF) videophone sequences is presented. Perceptually weighted cost/gain controlled motion compensation and quad-class DCT-based compression is applied without variable rate compression techniques and without adaptive buffering in order to maintain a fixed transmission rate, which can be adjusted to any required value. In this treatise we opted for a source coded rate of 11.36 kbps and the sensitivity-matched forward error correction (FEC) coded rate became 20.32 kbps. A partial forced update technique was invoked in order to keep transmitter and receiver aligned amongst hostile channel conditions. When using coherent pilot symbol assisted 16-level quadrature amplitude modulation (16-PSAQAM), an overall signalling rate of 9 kBd was yielded. Over lower quality channels 4QAM had to be invoked, which required twice as many time slots to accommodate the resulting 18 kBd stream. Over the best Gaussian and worst Rayleigh channels signal-to-noise ratio (SNR) values in the range of 7 to 20 dB were needed for these modems in order to maintain near-unimpaired image quality. In a bandwidth of 200 kHz, similarly to the GSM speech channel, 16 and 8 videophone users can be supported, when using the 16QAM and 4QAM systems, respectively.

1. INTRODUCTION

In recent years there has been an increased research activity in the field of videophony [1, 2, 3], in particular for mobile channels. In this treatise we propose a fixed-rate videophone codec, which can adjust its coding rate in order to accommodate its stream in a conventional speech channel, such as for example that of the Pan-European GSM system. In order to contrive such a GSM-like vidophone scheme the source rate was fixed at 11.36 kbps, and this stream was then transmitted using a transceiver, which can configure itself as a robust but less bandwidth efficient scheme or can double its bandwidth efficiency at the cost of requiring better channel conditions.

Section 2 outlines the design of the programmable video source codec, Section 3 investigates the bit sensitivities of the video source codec, while Section 4 details the design of the proposed re-configurable transceiver. The system's performance is characterised in Section 5, before offering some conclusions in Section 6.

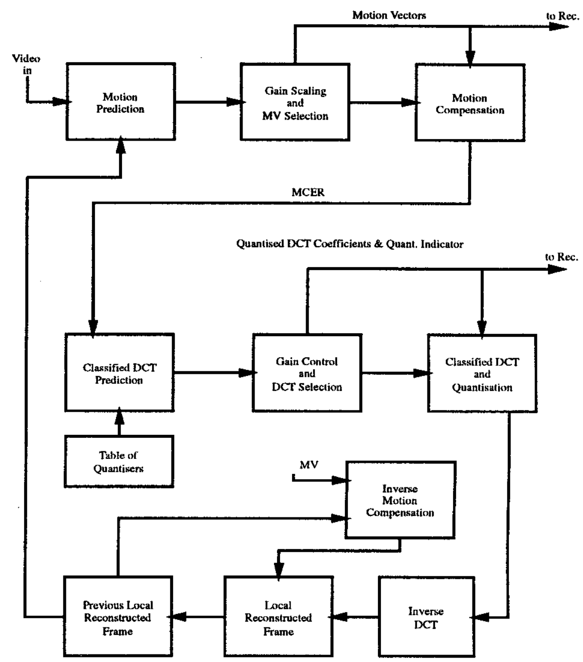

Figure 1: Encoder Schematic

2. VIDEOPHONE CODEC

Let us now focus our attention on the proposed video codec. In order to curtail error propagation across image frames the codec was designed to switch between intra- and inter-frame modes of operation. In the **intra-frame mode** the encoder transmits the coarsely quantised block averages for the current frame, which provides a low-resolution initial frame required for the operation of the inter-frame codec at both the commencement and during later stages of communications in order to prevent encoder/decoder misalignment.

The **inter-frame mode** of operation is based on a combination of gain-controlled motion compensation and gain-controlled DCT coding as seen in Figure 1. For the sake of communications convenience and simple networking our aim was to develop a fixed-rate codec which is able to dis-

0-7803-2431-5/95 $4.00 © 1995 IEEE

pense with an adaptive feed-back-driven bit-rate control buffer. Therefore a constant bit-rate source codec was required, which forced us to avoid using efficient variable-rate compaction algorithms, such as Huffman coding. This was achieved by fixing both the number of 8×8 blocks to be motion-compensated and those to be subjected to DCT to 30 out of 22×18=396. As mentioned, the selection of these blocks is based on a gain-controlled approach, which will be highlighted in the next Section.

Gain Controlled Motion Detection: At the commencement of the encoding procedure the motion compensation (MC) scheme determines a motion vector (MV) for each of the 8×8 blocks. The MC search window is fixed to 4×4 pels around the centre of each block. Before the actual motion compensation takes place the codec tentatively determines the potential benefit of the compensation in terms of motion compensated error energy reduction. In order to emphasize the subjectively more important eye and mouth region of the videophone images the potential gains for each motion compensated block are augmented by a factor of two in the centre of the screen. Then the codec selects the thirty blocks resulting in the highest scaled gain, and motion compensation is applied only to these blocks, whereas for all other so-called passive blocks the codec applies simple frame differencing.

Gain Controlled Quadruple-Class DCT: Pursuing a similar approach, gain control is also applied to the DCT-based compression. Every block is DCT transformed and quantised. Because of the non-stationary nature of the motion compensated error residual (MCER) the energy distribution characteristics of the DCT coefficients vary. Therefore four different sets of DCT quantisers are available, all of which are tentatively invoked in order to select the best set of quantisers resulting in the highest energy compaction gain. Ten bits are allocated for each quantiser, each of which are trained Max-Lloyd quantisers catering for a specific frequency-domain energy distribution class. Again, the energy compaction gain values are scaled to emphasise the eye and mouth region of the image and the DCT coefficients of the thirty highest-compression blocks are transmitted to the decoder.

Partial Forced Update: The disadvantage of interframe codecs is their vulnerability to channel errors. Every channel error results in a misalignment between the reconstructed frame buffer of the encoder and decoder. The errors accumulate and do not decay, unless a leakage-factor or a partial forced update (PFU) technique is employed. In our proposed codec in every frame 22 out of the 396 blocks, scattered over the entire frame, are periodically updated using the 4-bit quantised block means, which are partially overlayed on to the contents of the reconstructed frame buffer. The overlaying is performed such that the block's contents in the local buffer is weighted by 0.7 and superimposed on to the received block average, which is scaled by 0.3.

Bit Allocation Strategy: The bit allocation scheme was designed to deliver 1136 bits per frame, which begins with a 22 bit frame alignment word (FAW). This is necessary to assist the video decoder's operation in order resume synchronous operation after loss of frame synchronisation over hostile fading channels. The partial intra-frame update refreshes only 22 out of 396 blocks every frame. Therefore

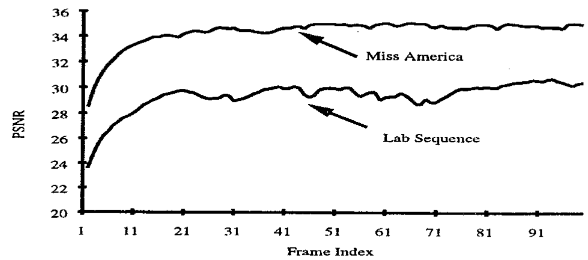

Figure 2: PSNR Codec Performance

every 18 frames or 1.8 seconds the update refreshes the same blocks. This periodicity is signalled to the decoder by transmitting the inverted FAW. A MV is stored using 13 bits, while 9 bits are allocated for the block index using the enumerative method and 4 bits for the X and Y displacement. The DCT-compressed blocks use a total of 21 bits, again 9 for the block index, 10 for the DCT coefficient quantisers, and 2 bits to indicate which quantiser has been applied. The total number of bits becomes $30 \cdot (13+21)+22 \cdot 4 + 22 + 6=1136$, where six dummy bits were added in order to obtain a total of 1136 bits suitable in terms of bit packing requirements for the specific forward error correction block codec used. The video codec's peak signal-to-noise ratio (PSNR) performance is portrayed in Figure 2 for the well-known 'Miss America' sequence and for a high-activity sequence referred to as the 'Lab sequence'.

3. SOURCE SENSITIVITY

In order to apply source-sensitivity matched protection the video bits were subjected to sensitivity analysis. In reference [3] we have consistently corrupted a single bit of a video coded frame and observed the image peak signal-to-noise ratio (PSNR) degradation inflicted. Repeating this method for all bits of a frame provided the required sensitivity figures and on this basis bits having different sensitivities can be assigned matching FEC codes. This technique, however, does not take adequate account of the phenomenon of error propagation across image frame boundaries. Therefore in this treatise we propose to use the method suggested in [4], where we corrupted each bit with a certain fixed error probability and observed the PSNR degradation in consecutive frames due to a single error event in the current frame. As an example, the PSNR degradation profile is shown for Bits 2 and 11 of the MV in Figure 3, where these bits were corrupted in all the 30 MVs.

In order to quantify the overall sensitivity of any specific bit we have integrated (summed) the PSNR degradations over the consecutive frames, where they have had a measurable effect and averaged these values for all the occurances of the corresponding bit errors. These results are shown in Figure 4 for the 13 MV bits, 21 DCT bits and 4 partial forced update bits.

4. SOURCE-MATCHED TRANSCEIVER

System Concept: The proposed system was designed for

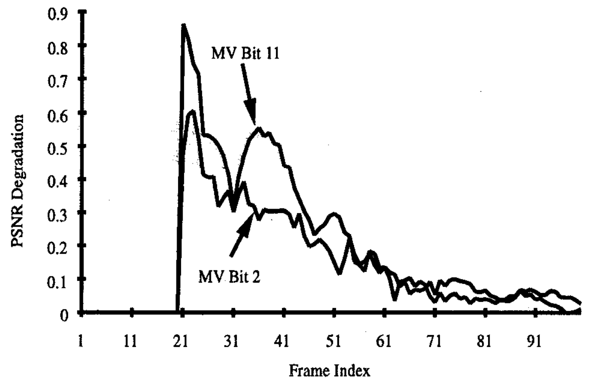

Figure 3: PSNR degradation profile for Bits 2 & 11 of the MV

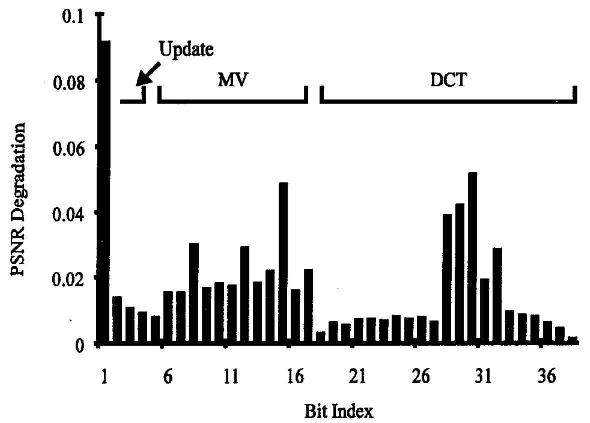

Figure 4: Integrated PSNR bit sensitivities

mobile packet video telephony and we experimented with two different modem schemes, namely 4-level and 16-level quadrature amplitude modulation (QAM). Our intention was to contrive a system, where the more benign propagation environment of indoors cells would benefit from the prevailing higher signal-to-noise ratio (SNR) by using 16QAM and thereby requiring only half the number of packets compared to 4QAM. When the portable station (PS) is handed over to an outdoors microcell or roams in a lower SNR region towards the edge of a cell, the base station (BS) instructs the PS to lower its number of modulation levels to 4 in order to maintain an adequate robustness under lower SNR conditions.

Sensitivity-matched Modulation: Best robustness against channel errors is achieved, if sensitivity-matched forward error correction coding is used. In our proposed videophone scheme we will exploit that 16-level pilot symbol assisted quadrature amplitude modulation [5] (16-PSAQAM) provides two independent 2-bit subchannels having different bit error rates (BER). Specifically, the BER of the higher integrity C1 subchannel is a factor 2-3 times lower than that of the lower quality C2 subchannel. Both subchannels support the transmission of two bits per symbol. This implies that the 16-PSAQAM scheme inherently caters for sensitivity-matched protection, which can be fine-tuned using appropriate FEC codes to match the source requirements. This property is unfortunately not retained by the 4QAM scheme, but the required different protection for the source coded bits can be ensured using appropriately matched channel codecs, if the associated higher system complexity is acceptable.

Source Sensitivity: In order to find the appropriate FEC code for our video codec its output stream was split in two equal sensitivity classes, Class One and Two according to our findings in Figure 4. Then the PSNR degradation of both classes was evaluated for a range of BER values. These results showed that a factor two lower BER was required by the Class One bits than by the Class Two bits, in order to maintain similar PSNR degradations in the range of 1-2 dB. This conveniently coincided with the integrity ratio of the C1 and C2 subchannels of our 16-PSAQAM modem. Hence we can apply the same FEC protection to both the Class One and Two source bits and direct the Class One bits to the C1 16-PSAQAM subchannel, while the Class Two bits to the C2 subchannel.

Forward Error Correction: Both convolutional and block codes can be successfully used over mobile radio links, but in our proposed scheme we have favoured a binary Bose-Chaudhuri-Hocquenghem (BCH) code. BCH codes combine a good burst error correction capability with reliable error detection, a facility useful to invoke image post-enhancement and to control handovers between traffic cells. The preferred $R=71/127 \approx 0.56$-rate BCH(127,71,9) code can correct 9 errors in a block of 127 bits, a correction capability of about 7.1 %. The number of channel coded bits per image frame becomes $1136 \times 127/71 = 2032$, while the bit rate is 20.32 kbps at an image frame rate of 10 frames/s.

Transmission Format: The transmission packets are constructed using one Class One BCH(127,71,9) code, one Class Two BCH(127,71,9) code and a third one is allocated to the packet header, yielding a total of 381 bits per packet. In case of 16QAM these are represented by 96 symbols and after adding 11 pilot symbols using a pilot spacing of $P = 10$ as well as 4 ramp symbols to ensure smooth power amplifier turn on/off the resulting 111-symbol packets are transmitted over the radio channel. Eight such packets represent a whole image frame and hence the signalling rate becomes 111 symb/12.5 ms $\approx$ 9 kBd. When using a time division multiple access (TDMA) channel bandwidth of 200 kHz, such as in the Pan-European second generation mobile radio system known as GSM and a modulation excess bandwidth of 38.8 %, the signalling rate becomes 144 kBd. This allows us to accommodate 144/9=16 users, which coincides with the number of so-called half-rate speech users supported by the GSM system.

When the prevailing channel SNR does not allow 16QAM communications, 4QAM must be invoked. In this case the 381-bit packets are represented by 191 2-bit symbols and after adding 20 pilot symbols and 4 ramp symbols the packet-length becomes 225 symb/12.5 ms, yielding a signalling rate of 18 kBd. In this case the number of videophone users supported becomes 8, as in the full-rate GSM speech channel. The system also facilitates mixed-mode operation, where 4QAM users must reserve two slots in each 12.5 ms TDMA

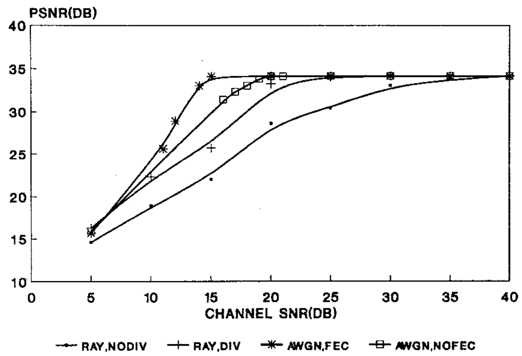

Figure 5: 16QAM PSNR versus channel SNR performance over various channels

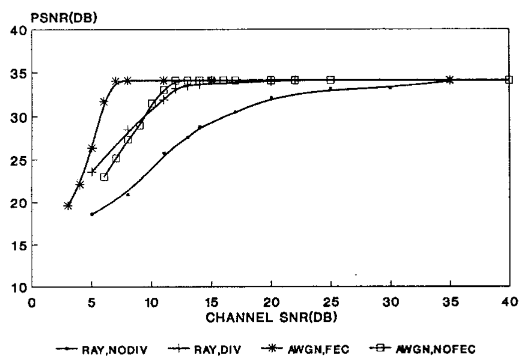

Figure 6: 4QAM PSNR versus channel SNR performance over various channels

frame towards the fringes of the cell, while in the central section of the cell 16QAM users will only require one slot per frame in order to maximize the number of users supported. Assuming an equal proportion of 4QAM and 16QAM users the average number of users per carrier becomes 12. The equivalent user bandwidth of the 4QAM PSs is 200 kHz/8=25 kHz, while that of the 16QAM users is 200 kHz/16=12.5 kHz.

5. SYSTEM PERFORMANCE

The video PSNR versus channel SNR performance of our videophone schemes is shown in Figures 5 and 6 for the 16QAM and 4QAM systems, respectively. The signalling rate was 144 kBd, the propagation frequency was 1.8 GHz and the vehicular speed was 30 mph. The system's performance was evaluated for the best-case additive white Gaussian noise (AWGN) channel and the worst-case Rayleigh (RAY) channel both with and without diversity (DIV). Observe in the Figure that the AWGN performance was evaluated also without forward error correction (FEC) coding in order to indicate the expected performance in a conventional tethered AWGN environment.

As expected, best performance is achieved over AWGN channels with FEC, requiring a channel SNR of about 15 and 7 dB in case of the 16QAM and 4QAM modems respectively, for an unimpaired image quality, which was associated with a PSNR value of about 35 dB. Without FEC coding over AWGN channels these SNR values must be increased to about 20 and 12 dB, respectively. The system performance over Rayleigh channels with second order diversity is similar to the AWGN, NOFEC scenario, while without diversity SNRs of about 30 and 20 dB were needed for near-unimpaired PSNR performance.

6. CONCLUSIONS

A highly bandwidth efficient, fixed-rate mobile videophone transceiver has been presented. The video source rate can be fixed to any arbitrary value in order be able to accommodate the videophone signal by conventional speech channels. In this treatise a source rate of 11.36 kbps was used, which is similar to the 13 kbps speech rate of the GSM system. After BCH(127,71,9) coding the channel rate becomes 20.32 kbps. When using an adaptive transceiver, which can invoke 16QAM and 4QAM depending on the channel conditions experienced, the signalling rate becomes 9 and 18 kBd, respectively. Accordingly, 16 or 8 videophone users can be accommodated in the GSM bandwidth of 200 kHz, which implies user bandwidths of 12.5 and 25 kHz, respectively. Over line-of-sight AWGN channels SNR values of about 15 and 7 dB are required to maintain unimpaired PSNR values of about 35 dB, which is increased to around 20 and 12 dB for the diversity-assisted Rayleigh scenario.

7. ACKNOWLEDGEMENT

The financial support of the EPSRC, UK in the framework of the Research Grant GR/J46845 is gratefully acknowledged.

8. REFERENCES

[1] IEEE Transactions on Circuits and Systems for Video Technology, Special issue on Very Low Bit Rate Video Coding, June 1994, Vol 4, No. 3, pp 213-357

[2] **L. Hanzo et al**: A Mobile Speech, Video and Data Transceiver Scheme, *Proc. of VTC'94, June 8-10, 1994, Stockholm, Sweden*, pp 452-456

[3] **R. Stedman, H. Gharavi, L. Hanzo, R. Steele,** Transmission of Subband-Coded Images via Mobile Channels, *IEEE Tr. on Circuits and Systems for Video Technology,* Feb 1993; Vol.3, No. 1, pp 15-27

[4] **L. Hanzo, J.P. Woodard** An intelligent cordless voice terminal for indoors communications, submitted to *IEEE Tr. on Veh. Technology,* 1994

[5] **W.T. Webb, L. Hanzo** Modern Quadrature Amplitude Modulation: Principles and Applications for Wireless Communications, *IEEE Press-Pentech Press*, 1994