WIRELESS VIDEOPHONE SCHEMES

J. Streit, L. Hanzo

Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO17 1BJ, UK.
Tel.: +44 703 593 125, Fax: +44 703 593 045
Email: jss@ecs.soton.ac.uk, lh@ecs.soton.ac.uk, http://www.ecs.soton.ac.uk

ABSTRACT

A range of 5-11.36 kbps videophone codecs are proposed and the 11.36 kbps Codec 1, the 8.52 kbps Codec 2 and the 8 kbps Codec 2a are embedded in the intelligent re-configurable Systems 1-3. After sensitivity-matched binary Bose-Chaudhuri-Hocquenghem (BCH) forward error correction (FEC) coding the data rate associated with Codec 1 and Codec 2a became 20.32 kbps, while that of Codec 2 was 15.24 kbps. When using Codec 1 in System 1 and coherent pilot symbol assisted 16-level quadrature amplitude modulation (16-PSAQAM), an overall signalling rate of 9 kBd was yielded. Over lower quality channels the 4QAM mode of operation had to be invoked, which required twice as many time slots to accommodate the resulting 18 kBd stream. In a bandwidth of 200 kHz, similarly to the Pan-European GSM mobile radio system's speech channel, using Systems 1 and 3 for example, 16 and 8 videophone users can be supported in the 16QAM and 4QAM modes, respectively. All system features are summarised in Table 3.

1. INTRODUCTION AND MOTIVATION

In recent years there has been an increased research activity in the field of very low rate video coding [1], in particular for mobile channels. In this treatise we set out to contrive a range of programmable constant-rate videophone codecs, which can adjust their coding rate in order to accommodate their stream in a conventional speech channel, such as for example that of the Pan-European GSM system [4]. The schematic of the proposed transceiver is portrayed in Figure 1. Section 2 outlines the design of a variety of programmable, but fixed-rate video source codecs, which is followed by the description of the source-matched transceiver in Section 3. The system's performance is characterised in Section 4.

2. VIDEOPHONE CODECS

2.1. Codec 1

The proposed 11.36 kbps discrete cosine transform [3] (DCT) based Codec 1 was designed for 176 × 144 pixels Quarter Common Intermediate Format (QCIF) images scanned at 10 frames/s. The required fixed low bit rate was achieved

ICUPC'95, 6-10 NOV. 1995, TOKYO, JAPAN

0-7803-2955-4/95/$4.00 ©IEEE

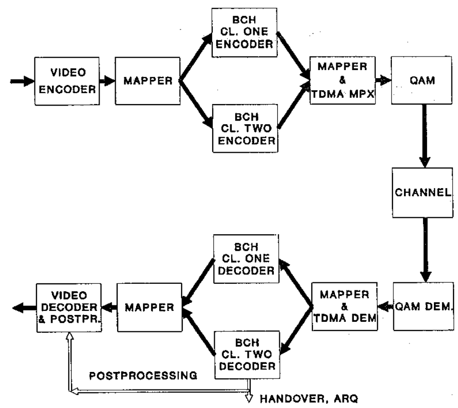

Figure 1: System's Schematic

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. The selection of these blocks is based on a gain-controlled approach, which will be highlighted next.

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.

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. 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 trained Max-Lloyd DCT quantisers are available. All four bit allocation schemes 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. 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 (PFU): 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. The bit-rate contribution of this PFU process is a moderate 22 $\times$4=88 bits per QCIF frame and it refreshes about 5.6 % of each frame.

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 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, where 9 bits are required to identify one of the 396 the block indexes using the enumerative method and 4 bits for encoding the 16 possible combinations of the X and Y displacements. The 8$\times$8 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 of the four 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. For 'Miss America' an average PSNR of about 33 dB was maintained, which was associated with pleasant videophone quality. [1]. The bit allocation scheme is summarised in Table 1. In our further discourse we will refer to the above scheme as Codec 1. Finally, the video bits were subjected to rigorous bit sensitivity analysis and a twin-class source-sensitivity matched error protection schemes was designed, which will be described later.

2.2. Codec 2

During our investigations we found that the motion activity table and the table of DCT-active blocks were potentially amenable to further data compression using run length coding (RLC). Therefore we set out to contrive a range of run length coded video codecs with bit rates as low as 5, 8 and 10 kbps, which we refer to as Codec 2. These tables can

[1] The MA sequence encoded at 11.36 kbps can be viewed under the address www: http://rice.ecs.soton.ac.uk

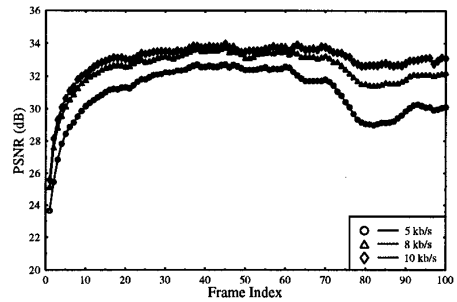

Figure 2: PSNR versus frame index performance of Codec 2 for the 'Miss America' sequence

be compressed using the elements of a two stage quad tree (QT) as follows.

First the 396-entry activity table containing the binary flags is grouped in 2 $\times$ 2 blocks and a four bit symbol is allocated to those blocks which contain at least one active flag. These four-bit symbols are then run length encoded and transmitted to the decoder. This concept requires a second active table containing 396 / 4 = 99 flags in order to determine which of the two by two blocks contain active vectors. Three consecutive flags in this table are packetised to a symbol an then run length encoded. As a result, a typical 396-bit active/passive table containing 30 active flags can be compressed to less than 150 bits. The motion vectors do not lend themselves to run length encoding.

If at this stage of the encoding process the number of bits allocated to the compressed motion- and DCT-activity tables as well as to the active MVs exceeds half of the total number of available bits/frame, some of the blocks satisfying the initial motion-active criterion will be relegated to the motion-passive class. This process takes account of the subjective importance of various blocks and and does not ignore motion-active blocks in the central eye and lip regions of the image, while relegating those, which are closer to the fringes of the frame. The DCT blocks are handled using a similar procedure. Depending on the actual fixed-length transmission burst and the free buffer space, a number of active DCT blocks is chosen and the corresponding compressed tables are determined. If the total bit count overspills the transmission burst or if there are too many bits left unused, a different number of active blocks is estimated and new tables are determined.

The PSNR versus frame index performance of a 5, 8 and 10 kbps RLC scheme is shown for the 'Miss America' sequence in Figure 2 and the average results are summarised is Table 2. Based on these results, in the run length coded System 2 we have opted for an **8.52 kbps** implementation of **Codec 2**, generating 852 bits per frame and maintaining an average PSNR of about 33.3 dB for the MA sequence. We also note that in some of the proposed systems an **8 kbps** reduced-rate version of Codec 2 will be invoked, which we refer to as **Codec 2a**. Again, the source sensitivity-matched bit protection scheme will be described at a later stage.

FAW	PFU	MV Index	MV	DCT Index	DCT	Padding	Total
22	22×4	30×9	30×4	30×9	30×12	6	1136

Table 1: Bit Allocation Table

Sequence	'Miss America'	'Lab'
5 kb/s	30.26 dB	21.87 dB
8 kb/s	33.29 dB	24.34 dB
10 kb/s	33.52 dB	26.91 dB

Table 2: Average PSNR performance of Codec 2 for the 'Miss America' and 'Lab' sequences

3. SOURCE-MATCHED TRANSCEIVER

3.1. System 1

System Concept: The system's schematic is portrayed in Figure 1. In the proposed system the more benign propagation environment of indoors cells would benefit from the prevailing higher signal-to-noise ratio (SNR) by using bandwidth efficient 16-level quadrature amplitude modulation [2] (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. Let us now focus our attention on specific details of System 1.

Transmission Format: In reference [2] we have shown that 16QAM can be viewed as a twin-channel (C1 and C2) modulation scheme, since two of its four bits have about two or three times lower bit error rate (BER) than the other two. This property can be exploited to provide source sensitivity matched protection for the more vulnerable Class One source bits by transmitting them over the higher integrity C1 16QAM subchannel and relegating the less sensitive Class Two video bits to the C2 16QAM subchannel. The transmission packets are constructed using one Class One BCH(127,71,9) code [4], one Class Two BCH(127,71,9) code and a stronger BCH(127,50,13) 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 ramping 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 [4].

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 by System 1 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 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. The above-mentioned features of the 16QAM/4QAM System 1 along with the characteristics of a range of other systems about to be introduced in the next Section are summarised in Table 3.

3.2. System 2

The lower-rate Codec 2 became more vulnerable against transmission errors than Codec 1 and their effect is particularly objectionable, if the run length coded activity table bits are corrupted. Therefore in System 2, which was designed to incorporate Codec 2, the more sensitive run length coded activity table bits are protected by the powerful binary Bose-Chaudhuri-Hocquenghem BCH(127,50,13) codec, while the less vulnerable remaining bits by the weaker BCH-(127,92,5) code. Note that the overall coding rate of $R = (50+92)/(127+127) \approx 0.63$ is identical to that of System 1, but the RL-coded Class One bits are more strongly protected. At a fixed coding rate this inevitably assumes a weaker code for the protection of the less vulnerable Class Two bits. The 852 bits/100ms video frame is encoded using six pairs of such BCH code words, yielding a total of $6 \cdot 254=1524$ bits, which is equivalent to a bit rate of 15.24 kbps.

As in System 1, the more vulnerable run length and BCH(127,50,13) coded Class One bits are then transmitted over the higher integrity C1 16QAM subchannel. The less sensitive BCH(127,92,5) coded Class Two DCT coefficient bits are conveyed using the lower-integrity C2 16QAM subchannel. This arrangement is favoured in order to further emphasize the integrity differences of the BCH codecs used, which is necessitated by the integrity requirements of the video bits.

The transmission burst is constructed by adding an additional BCH(127,50,13) code word for the packet header and the resulting 381 bits are again converted to 96 16QAM symbols, and pilot as well as ramp symbols are added. In System 2 six such packets represent a video frame, hence the single-user signalling rate becomes 666 symb/100 ms, which corresponds to 6.66 kBd. This allows us to accommodate now Integer[144 kBd/6.66]=21 such users, if no time slots are reserved for packet re-transmissions. This number will

Feature	System 1	System 2	System 3
Video Codec	Codec 1	Codec 2	Codec 2a
Video rate (kbps)	11.36	8.52	8
Frame Rate (fr/s)	10	10	10
C1 FEC C2 FEC Header FEC	BCH(127,71,9) BCH(127,71,9) BCH(127,50,13)	BCH(127,50,13) BCH(127,92,5) BCH(127,50,13)	(BCH(127,50,13) (BCH(127,50,13) (BCH(127,50,13)
FEC-coded Rate (kbps)	20.36	15.24	20.36
Modem	4/16-PSAQAM	4/16-PSAQAM	4/16-PSAQAM
ARQ	None	Cl. One	None
User Signal. Rate (kBd)	18 or 9	6.66	18 or 9
System Signal. Rate (kBd)	144	144	144
System Bandwidth (kHz)	200	200	200
No. of Users	8 or 16	(21-2)=19	8 or 16
Eff. User Bandwidth (kHz)	25 or 12.5	10.5	25 or 12.5
Min. AWGN SNR (dB)	7 or 15	15	8 or 12
Min. Rayleigh SNR (dB)	15 or 20	25	15 or 16

Table 3: Summary of System Features

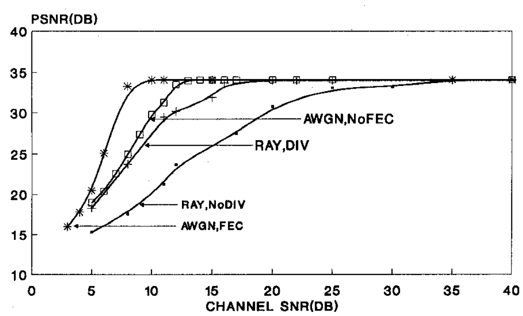

Figure 3: 4QAM PSNR versus channel SNR performance of System 1 in its 18 kBd mode of operation over various channels

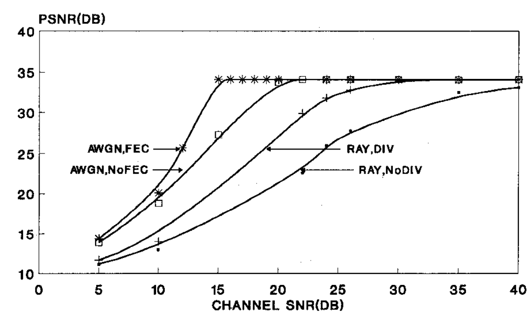

Figure 4: 16QAM PSNR versus channel SNR performance of System 1 in its 9 kBd mode of operation over various channels

have to be reduced in order to accommodate ARQs.

In System 2 when the more powerful BCH codec conveying the more sensitive run-length coded Class One bits over the C1 16QAM subchannel is overloaded by channel errors, we re-transmit these bits only, using robust 4QAM. Explicitly, for the first transmission attempt (TX1) we use contention-free Time Division Multiple Access (TDMA). If an ARQ-request occurs, the re-transmitted packets will have to contend for a number of earmarked time slots, which are allocated on a firts come, first served basis. The intelligent base station (BS) detects these events of packet corruption and instructs the portable stations (PS) to re-transmit their packets during the slots dedicated to ARQ-packets. Reserving slots for ARQ-packets reduces the number of video users supported depending on the prevailing channel conditions, as we will show in the Results Section, Section 4.

Although the probability of erroneous packets can be reduced by allowing repeated re-transmissions, there is a clear trade-off between the number of maximum transmission attempts and the BCH-coded frame error rate (FER). In order to limit the number of slots required for ARQ-attempts, which potentially reduce the number of video users supported, in System we invoke ARQ only, if the more sensitive run-length coded Class One bits transmitted via the C1 16-PSAQAM channel and protected by the BCH(127,50,13) codec are corrupted. Furthermore, we re-transmit only the Class One bits, but in order to insure a high success rate, we use 4-PSAQAM, which is more robust than 16-PSAQAM. Since only half of the information bits are re-transmitted, they can be accommodated within the same slot interval and same bandwidth, as the full packet. If there are only C2 bit errors in the packet, it is not re-transmitted, which implies that typically there will be residual Class Two errors. In order to limit the number slots dedicated to re-transmissions we limited the number of transmission attempts to three, which implies that a minimum of two slots per frame must be reserved for ARQ. In order to maintain a low system complexity we dispense with any contention mechanism and allocate two time slots to that particular user, whose packet was first corrupted within the TDMA frame. Further users cannot therefore invoke ARQ, since there are no more unallocated slots. A further advantage is that in possession

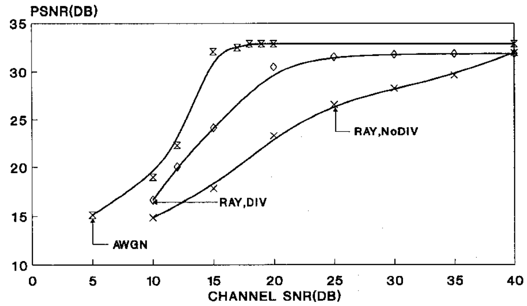

Figure 5: PSNR versus channel SNR performance of the 6.6 kBd System 2 over various channels

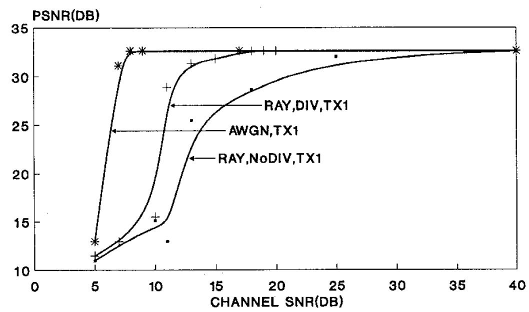

Figure 6: PSNR versus channel SNR performance of the 18 kBd 4QAM mode of System 3 over various channels

of three copies of the transmitted packet majority decisions can be invoked, if all three packets became corrupted. The basic features of System 2 designed to accommodate Codec 2 are also summarised in Table 3.

3.3. System 3

System 3 employs a run-length coded source compression scheme similar to Codec 2, but with a slightly reduced bit rate of 8kbps or 800 bits per frame, which we refer to as **Codec 2a**. In this system the Class One and Two bits were protected by the more powerful BCH(127,50,13) code instead of the BCH(127,71,9) scheme, but the modem was identical to that in Systems 1 and 2. The slightly reduced video rate of 8kbps was imposed in order to be able to accommodate the BCH(127,50,13) code in both 16QAM subchannels, while maintaining the same 20.36 kbps overall rate, as System 1. Clearly, System 3 will allow us to assess, whether it is a worthwhile complexity investment to introduce run-length coding in Codec 1 in order to reduce the source bit rate and whether the increased error sensitivity of Codec 2a can be compensated for by accommodating the more complex and more powerful BCH(127,50,13) codec. Again, these system features are summarised in Table 3. Having designed the video transceivers their performance results are presented in the next Section.

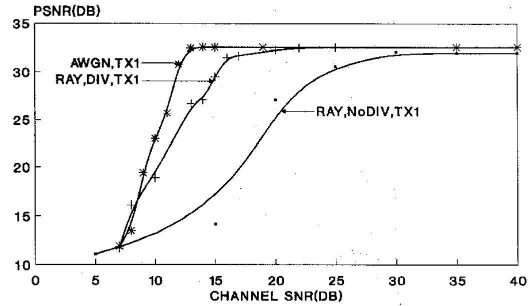

Figure 7: PSNR versus channel SNR performance of the 9 kBd 16QAM mode of System 3 over various channels

4. SYSTEM PERFORMANCE

In our experiments the signalling rate was 144 kBd, while the propagation frequency and the vehicular speed were 1.8 GHz and 30 mph, respectively. For pedestrian speeds the fading envelope fluctuates less dramatically and hence our experimental conditions constitute a GSM-like urban worst-case scenario.

The video PSNR versus channel SNR (ChSNR) performance of System 1 is shown in Figures 3 and 4 for the 4QAM and 16QAM modes of operation, respectively. The channel SNR values required for near-unimpaired PSNR video performance are tabulated for all systems in Table 3. The PSNR versus ChSNR performance of System 2 is characterised by Figure 5. Observe that the 6.6 kBd System 2 has a lower robustness than the 9 kBd System 1, since its behaviour is predetermined by the initially transmitted Class Two video bits, which were protected by the weaker BCH(127,92,5) code and were never re-transmitted. Furthermore, System 2 is inherently more complex than System 1 and only marginally more bandwidth efficient. Therefore we preferred System 1 to System 2. Lastly, the PSNR versus ChSNR performance of System 3 is displayed in Figures 6 and 7.

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

[3] **A.K. Jain**: Fundamentals of Digital Image Processing, Prentice-Hall, 1989

[4] R. Steele (Ed.): Mobile radio communications, IEEE Press-Pentech Press, 1994 Channel Coding, pp 347-488,