VECTOR-QUANTISED CORDLESS VIDEOPHONE TRANSCEIVERS

J. Streit, L. Hanzo

Dept. of Electr. and Comp. Sc., Univ. of Southampton, SO17 1BJ, UK.
Tel: +44 1703 593 125, Fax: +44 1703 593 045
Email: jss@ecs.soton.ac.uk, lh@ecs.soton.ac.uk
http://www-mobile.ecs.soton.ac.uk

ABSTRACT

A variety of adaptively re-configurable wireless videophone transceivers are proposed and their video quality, bit rate, robustness and complexity issues are analysed. A suite of fixed but arbitrarily programmable low-rate, perceptually weighted vector quantised (VQ) codecs with and without run-length compression (RLC) are contrived for quarter common intermediate format (QCIF) videophone sequences. The 11.36 kbps Codec 1 is BCH(127,71,9) coded to a rate of 20.32 kbps and this arrangement is comparatively studied along with the 8 kbps Codec 2 and BCH(127,50,13) scheme, which has the same 20.32 kbps overall rate. The source-sensitivity matched Systems 1-6 characterised in Table 3 were contrived to comparatively study the range of system design options. For example, 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. The robustness of Systems 2, 3, 4 and 6 was increased using Automatic Repeat Requests (ARQ), inevitably reducing the number of users supported, which was between 6 and 16. In a bandwidth of 200 kHz, similarly to the Pan-European GSM mobile radio system's speech channel, using Systems 1, 3, 4 or 5 for example, 16 and 8 videophone users can be supported in the 16QAM and 4QAM modes, respectively. The basic system characteristics are highlighted in Table 3.

1. BACKGROUND

Motivated by the the rapid emergence of mobile video telephony as a major research area [1, 2], in this treatise we set out to contrive and study a range of vector-quantised (VQ) wireless videophone schemes. The structure of the reconfigurable transceiver is shown in Figure 1, where the VQ video stream is mapped in two different protection classes, namely class one and two. The VQ bits are binary Bose-Chaudhuri-Hochquenghem (BCH) coded [4], before they are multiplexed to a Time Division Multiple Access (TDMA) slot and transmitted using Quadrature Amplitude Modulation (QAM) [3].

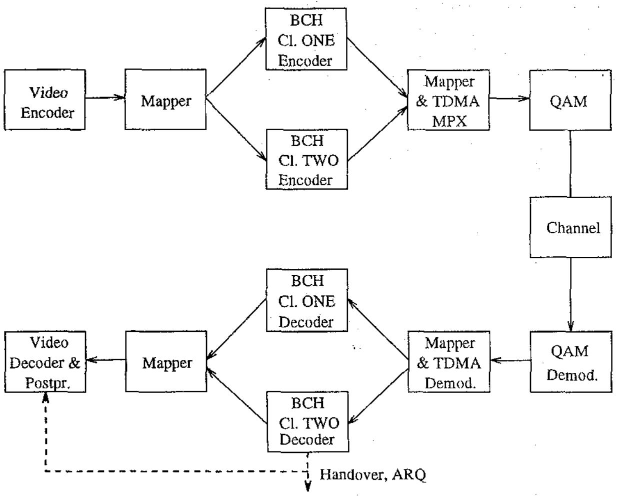

Figure 1: System's Schematic

Section 2 highlights the design of the 11.36/8 kbps VQ video codecs, while Section 3 discusses modulation, forward error correction (FEC) coding and system performance issues, before concluding in Section 4.

2. VECTOR-QUANTISED VIDEO CODECS

2.1. The Codebook Design

The codebook design is a crucial issue for every VQ codec. In our experiments we calculated the Displaced Frame Difference (DFD) signal for various input sequences while only allowing for 35 active motion vectors out of the 396 8×8 blocks of a Quarter Common Intermediate Format (QCIF) frame to be generated. This presumes that the VQ will be used in very low bitrate image coding, where an active/passive block classification is necessary. The DFD energy was analysed and the 35 blocks containing the highest energy were copied into the training sequence.

Initially the codebook was designed using the so-called 'pruning' method [6], but its preformance was inadequate. In a second attempt, we made use of the 'pairwise nearest

---

THIS TREATISE IS COMPLEMENTED BY A DEMONSTRATION PACKAGE PORTRAYING VIDEO SEQUENCES AT VARIOUS BIT RATES, WHICH IS DOWN-LOADABLE FROM HTTP://WWW-MOBILE.ECS.SOTON.AC.UK
GLOBECOM'96, 18-22 NOV. 1996, LONDON, UK

0-7803-3336-5/96 $5.00 © 1996 IEEE

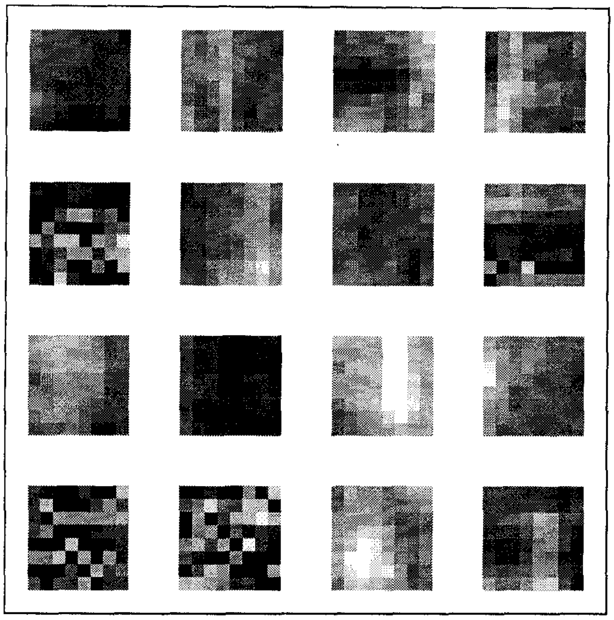

Figure 2: Enhanced sample codebook with sixteen 8 × 8 vectors

neighbour' (PNN) algorithm [6]. This approach shrinks a given codebook step by step until a desired size is achieved. Initially each vector is assigned to a separate cluster. Then for two candidate clusters the distortion penalty incurred by merging these two clusters is determined. This is carried out for all possible pairs of clusters and finally the pair with the minimum distortion penalty is merged to a single cluster. This process is continued until the codebook is shrunk to the desired size. The algorithm's complexity increases drastically with the codebook size and this technique became impractical for our large training set. Hence we simplified this algorithm by limiting the number of tentative cluster combinations as follows. Instead of attempting to merge each possible pair of clusters, where the number of combinations exhibits a quadratic expansion with increasing codebook size, in our approach a single cluster is preselected and combined with all the others.

Following this technique the codebook length was initially reduced and then the full PNN algorithm was invoked in order to create a range of codebooks having sizes between 4 and 1024. As an example, our sixteen-entry codebook constituted by 8 × 8 vectors is depicted in Figure 2, where the vector components were shifted by 127 and multiplied by 30 in order to visually emphasize their differences.

2.2. VQ Codec Design

The schematic of the proposed image codec is displayed in Figure 3. A constant bit-rate source codec was required, in order to be able accommodate its bit stream in conventional mobile radio speech channels, such as for example that of the GSM system [5]. Hence we fixed both the number of 8×8 blocks to be motion-compensated and those to be vector quantisatised to a value depending on the required bit rate.

Intra-frame Mode: Since the first frame used in motion compensation is unknown to the decoder, the 4-bit encoded block averages are transmitted first in the intra-frame mode, which are also used by the local decoder. In order to mitigate the effects of transmission errors Partial Forced Update (PFU) of both the encoder's and decoder's reconstructed buffers is also carried out on a similar basis, using the 4-bit encoded block averages of a certain number of blocks. The number of the PFU blocks per frame is also bit rate dependent and it is automatically determined by our programmable codec. For example, in our 11.36 kbps prototype 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 ×4=88 bits per QCIF frame and it refreshes about 5.6 % of each frame.

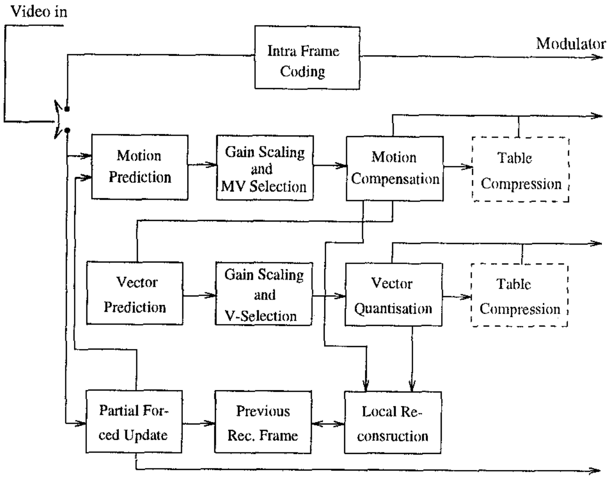

Figure 3: Basic Schematic of the VQ

Gain Controlled Motion Compensation: Initially 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. In our cost-gain controlled approach the codec tentatively determines the achievable gain of the compensation in terms of displaced frame difference (DFD) energy reduction. The gains in the most important eye and mouth region are weighted by a factor of two. Then the codec identifies the required number of blocks resulting in the highest scaled DFD gain, and motion compensation is applied only to these blocks. For the remaining motion-passive blocks frame differencing is employed. Observe in Figure 3 that an optional 'Table Compression' algorithm can be employed in order to further improve the codec's bandwidth efficiency, which will be described during our further discourse.

Vector Quantisation: The DFD signal is then passed on to the VQ scheme. A second gain- and bitrate controlled algorithm decides for which of the blocks the VQ provides a high coding gain, ie which of the blocks are deemed to be active and these indices are then passed on to the receiver. Explicitly, the active VQ block indices are stored and transmitted after optional activity table compression, as seen in Figure 3, which will be described at a later stage.

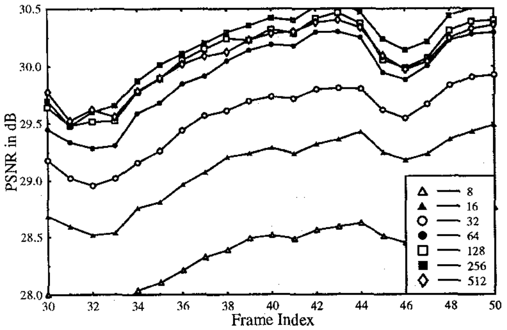

Figure 4: PSNR performance for the 'Lab Sequence' and various codebook sizes at a constant bit rate

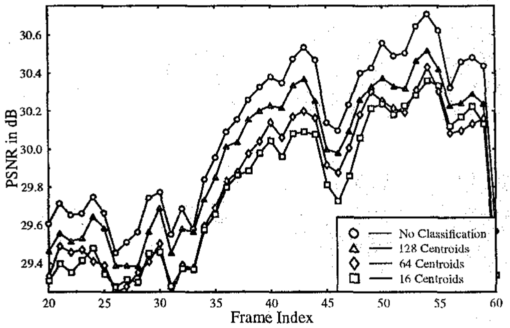

Figure 5: PSNR versus frame index performance of classified VQ codecs when using the 'Lab Sequence'

The encoded frame is locally reconstructed and fed back to the 'Previous Reconstructed Frame Buffer' in order to assist in the next stage of motion compensation.

For every block the VQ searches through the entire codebook and compares the match by summing the squared differences for each pixel within the two blocks under consideration. Therefore 191 Floating Point Operations (Flops) are necessary to perform one block match. In case of a 128-entry codebook for an image frame a total computational complexity of 396 blocks × 128 codebook entries × 191 Flop = 9.6 Mflops is incurred. Similarly, a 256-entry codebook implies a complexity of 19.3 Mflops.

We opted for a fixed but programmable bit rate scheme and varied the codebook size between four and 1024. As Figure 4 reveals, best Peak Signal-to-Noise Ratio (PSNR) performance was achieved using codebook sizes in the range of 128 and 512. For these investigations we used a locally recorded high-activity head-and-shoulders videophone sequence, which we refer to as the 'Lab Sequence', since the well-known low-activity Miss America (MA) sequence was inadequate for evaluating the VQ performance. Observe that the 256-entry codebook results in the best PSNR performance. This corresponds to a VQ data rate of 0.14 bit per pixel or eight bit per 8×8 block. However, a codebook size of 128 is preferable, as it halves the codec's complexity without significantly reducing the quality.

The classified vector quantiser (CVQ) [6] is essentially a combination of a classifier and an ordinary VQ using a series of codebooks. The incoming block is classified into one of n classes and hence only the corresponding reduced-size codebook c_n is searched, thereby reducing the search complexity. In our approach a smaller codebook, derived from the same training sequence as the un-classified codebook C, can be viewed as a set of centroids for the codebook C. Hence, codebook C can be split into n codebooks c_n by assigning each vector in C to one of the centroids. The encoding procedure then consists of a two-stage VQ process and may be seen as a tree-structured VQ (TSVQ) [6]. The classifier is a VQ in its own right, using a codebook filled with the n centroids. Once the closest centroid has been found, the sub-codebook containing the associated vectors is selected for the second VQ step. We carried out various experiments, based on the optimum 256-entry codebook, with classifiers containing n=16, 64 and 128 centroids, as revealed in Figure 5. The PSNR performance loss due to using this sub-optimum two-stage approach is less than 0.3 dB. This is amazing, as the number of block comparisons was reduced from 256 for the standard VQ to about 64 for the CVQ. The complexity reduction cannot be exactly quantified, as the codebooks c_n do not necessarily contain N/n vectors.

2.3. Bit Allocation Strategy

Finally, in order to be able to accommodate the video encoded bit stream in a conventional mobile radio speech channel we contrived two VQ-based prototype codecs. Codec 1 generated a video source rate of 11.36 kbps or 1136 bits per frame and dispensed with the optional motion- and VQ-activity table compression shown in dashed lines in Figure 3. This ensured a higher innate robustness than in case of the run-length compressed 8 kbps Codec 2. The bit allocation scheme of both codecs is shown in Table 1, which will be detailed during our following discourse.

Both codecs are based on the classified VQ codec as it offers the best compromise in terms of quality, robustness and complexity. We used a centroid codebook size of 16 and an overall codebook size of 256 which lead to an overall codec complexity of around 15 Mflops when using the following active/passive classification. Initially the frame difference signal is computed for all 396 8×8 blocks and only a certain fraction of the high-energy frame difference blocks between 20-50 % is deemed active on this basis, which then undergo further processing as detailed below. This initial classification results in a significant complexity reduction at the price of a minor PSNR performance degradation.

Codec 1 did not take advantage of the optional run-length encoding for the active/passive tables for the sake of increased robustness. The codec output contains the frame alignment word (FAW), the PFU, the MVs and the VQ data. The 22 bit FAW is required to support the video decoder's operation in order to regain synchronous operation after loss of frame synchronisation. The partial intra-frame update refreshes 22 8 × 8-sized blocks out of the 396 blocks per 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. In order to comply with the bit rate requirement of 11.36 kbps, a total of 38 MVs are stored using 13 bits each, where 9 bits are required to identify one of the 396 indices using the enu-

Codec	FAW	PFU	MV Index	MV	VQ Index	VQ	Padding	Total
Codec 1	22	22×4	38×9	38×4	31×9	31×8	5	1136
Codec 2	22	22×4	-	< 500 (VLC)	-	VLC	VLC	800

Table 1: Bit Allocation Table

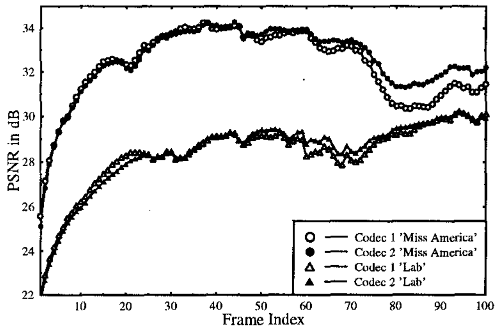

Figure 6: PSNR versus frame index performance of the 11.36 kbps Codec 1 and the 8 kbps Codec 2 for the 'Miss America' sequence

merative method and 4 bits for encoding the 16 possible combinations of the X and Y displacements. Each of the 31 active 8×8 VQ blocks uses a total of 17 bits, again 9 for the block index, 8 for the codebook index. The total number of bits becomes $22 + 22 \cdot 4 + 38 \cdot 13 + 31 \cdot 17 + 5 = 1136$, where five 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 6 for the well-known 'Miss America' (MA) sequence and for a high-activity sequence referred to as the 'Lab sequence' [1]. For 'Miss America' an average PSNR of about 33 dB was maintained, which was associated with pleasant videophone quality.

Again, the bit allocation scheme is summarised in Table 1 and the complexity of this codec is about 15 Mflops when applying the previously introduced active/passive block classification prior to further processing. Recall that this classification process determines the energy of every block prior to the MC or VQ encoding steps and only the 100 blocks associated with the largest energy values are subjected tentatively to full MC and VQ.

Codec 2 capitalises on the fact that the tables of active MV and active VQ-block indices retain some redundancy, which can be removed by Variable-Length Coding (VLC), as suggested by the dashed blocks of Figure 3. In order to be able to transmit all block averages in the intra-frame mode with a 4-bit resolution, as in Codec 1, while not exceeding the 800 bits/frame budget we fixed the intra frame block size to 12 × 12. Again, the full bit allocation scheme is portrayed in Table 1.

Codec / Sequence	'Miss America'	'Lab'
Codec 1 (11.36 kb/s)	32.56 dB	28.56 dB
Codec 2 (8 kb/s)	32.62 dB	28.60 dB

Table 2: Average PSNR performance of Codecs 1 and 2 for the 'Miss America' and 'Lab' sequences

However, in the motion-compensation (MC) we retained the block-size of 8×8 and the search window size of 4×4 around the centre of each block. This method of classifying the blocks as motion-active and motion-passive results in an active/passive table, which consists of a one bit flag for each of the 396 blocks, marking it as passive or active. These tables are 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 × 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 and 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-activity tables as well as to the active MVs exceeds half of the total number of bits available for the encoding of each frame, which is 800 in case of the 8kbps Codec 2, then some of the blocks satisfying the initial motion-active criterion will be relegated to the motion-passive class.

The VQ blocks are handled using a similar procedure. Depending on the required number of bits per frame and the free buffer space, an appropriate number of active VQ blocks is chosen and the corresponding compressed tables are determined. If the total number of bits 'overspills', or if there are too many bits left unused, a different number of active blocks is estimated and new tables are determined. The disadvantage of such a high compression scheme is the strongly increased vulnerability of the transmission burst. If a transmission error corrupts one of the RL encoded tables, it is likely that a codeword of a different length is generated and decoding becomes impossible. This will force the decoder to drop the entire frame.

The average PSNR performances of Codecs 1 and 2 are summarised for the MA sequence and for the higher activity 'Lab Sequence' in Table 2.

---

[1] The MA sequence encoded at various bit rates can be viewed under the WWW address http://rice.ecs.soton.ac.uk

Feature	System 1	System 2	System 3	System 4	System 5	System 6
Video Codec	Codec 1	Codec 2	Codec 1	Codec 2	Codec 2	Codec 1
Video rate (kbps)	11.36	8	11.36	8	8	11.36
Frame Rate (fr/s)	10	10	10	10	10	10
C1 FEC	BCH(127,71,9)	BCH(127,50,13)	BCH(127,71,9)	BCH(127,50,13)	BCH(127,50,13)	BCH(127,71,9)
C2 FEC	BCH(127,71,9)	BCH(127,92,5)	BCH(127,71,9)	BCH(127,50,13)	BCH(127,50,13)	BCH(127,71,9)
Header FEC	BCH(127,50,13)	BCH(127,50,13)	BCH(127,50,13)	BCH(127,50,13)	BCH(127,50,13)	BCH(127,50,13)
FEC-coded Rate (kbps)	20.32	20.32	20.32	20.32	20.32	20.32
Modem	4/16-PSAQAM	4/16-PSAQAM	4/16-PSAQAM	4/16-PSAQAM	4/16-PSAQAM	4/16-PSAQAM
ARQ	None	Cl. One	Cl. One & Two	Cl. One & Two	None	Cl. One
User Signal. Rate (kBd)	18 or 9	9	18 or 9	18 or 9	18 or 9	9
System Signal. Rate (kBd)	144	144	144	144	144	144
System Bandwidth (kHz)	200	200	200	200	200	200
No. of Users	8 or 16	(16-2)=14	6 or 14	6 or 14	8 or 16	(16-2)=14
Eff. User Bandwidth (kHz)	25 or 12.5	14.3	33.3 or 14.3	33.3 or 14.3	33.3 or 14.3	14.3
Min. AWGN SNR (dB) 4/16QAM	5/11	11	4.5/10.5	6/11	8/12	12
Min. Rayleigh SNR (dB) 4/16QAM	10/22	15	9/18	9/17	11/17	17

Table 3: Summary of System Features

3. TRANSMISSION ISSUES AND PERFORMANCE

The VQ codec's output bit stream was subjected to bit sensitivity analysis and the bits were sorted in two bit sensitivity classes. A twin-class binary Bose-Chaudhuri-Hocquenghem (BCH) coding [4], scheme was invoked and the bit stream was transmitted using bandwidth-efficient Quadrature Amplitude Modulation [3] (QAM). In this contribution we created a range of re-configurable source-sensitivity matched videophone transceivers, which have two modes of operation. Namely, a more robust but less bandwidth efficient 4-level Quadrature Amplitude Modulation (4QAM) [3] mode of operation, accommodating eight users in a bandwidth of 200 kHz, as in the GSM system [5] and a less robust but more bandwidth efficient 16QAM mode, supporting 16 users. For example, indoors cells exploit the prevailing higher signal-to-noise ratio (SNR) and signal-to-interference ratio (SIR) invoking 16QAM and thereby requiring only half as many packets as 4QAM. When the portable station (PS) is roaming in a lower-SNR outdoors cell, the intelligent base station (BS) re-configures the system to operate at 4QAM. Reference [3] suggested that the rectangular 16QAM constellation possesses two independent 2-bit subchannels having different bit error rates (BER), which naturally lend themselves to un-equal protection coded modulation. The BER of the lower integrity C2 subchannel was found a factor 2-3 times higher than that of the C1 subchannel. These BER differences can be adjusted using appropriate FEC codes to match arbitrary bit sensitivity requirements.

In order to explore the range of system design options, we contrived 6 different systems and in Systems 1, 3 and 6 we used the $R=71/127 \approx 0.56$-rate BCH(127,71,9) code in both 16QAM subchannels, which, in conjunction with the 11.36 kbps video Codec 1, resulted in $1136 \times 127/71 = 2032$ bits/frame and a bit rate of 20.32 kbps at an image frame rate of 10 frames/s. These system features are summarised in Table 3. In contrast, in Systems 2, 4 and 5 we employed the 8 kbps video Codec 2, which achieved this lower rate essentially due to invoking RL coding in order to compress the motion- and VQ-activity tables. Since erroneous RL-coded bits corrupt the entire video frame, their protection is crucial. Hence in Systems 2, 4 and 5 we decided to use the stronger BCH(127,50,13) code, which after FEC coding yielded the same 20.32 kbps bit rate, as the remaining systems.

The video PSNR versus channel SNR performance of Systems 1-6 was evaluated over both Additive White Gaussian Noise (AWGN) and Rayleigh fading channels in both their 4QAM and 16QAM modes of operation with and without 2nd order switch-diversity and Automatic Repeat Request (ARQ). The minimum required channel SNR, as well as the number of users supported was summarised in Table 3.

4. SUMMARY AND CONCLUSIONS

Due to lack of space specific details of the proposed systems have not been described, although their salient features have been summarised in Table 3. The video codecs proposed are programmable to any arbitrary transmission rate in order to host the videophone signal by conventional mobile radio speech channels, such as the Pan-European GSM system, the IS-54 or IS-95 systems as well as the Japanese digital cellular system.

5. ACKNOWLEDGEMENT

The financial support of the EPSRC, UK in the framework of the research contract GR/K74043 is gratefully acknowledged.

6. REFERENCES

[1] **J. Streit, L. Hanzo**: Vector-quantised low-rate cordless videophone systems, submitted to IEEE Tr. on Veh. Techn., 1995

[2] 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

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

[4] **K.H.H. Wong, L. Hanzo**: Channel Coding, pp 347-488, Chapter 4 in *R. Steele (Ed.) Mobile Radio Communications, IEEE Press-Pentech Press*, London, 1992

[5] **L. Hanzo, J. Stefanov**: The Pan-European Digital Cellular Mobile Radio System - known as GSM, Chapter 8, pp 677-773, in *R. Steele (Ed.) Mobile Radio Communications, Pentech Press, London. 1992*

[6] **A. Gersho, R.M. Gray**: Vector Quantization and Signal Compression, *Kluwer Academic Publishers*, 1992