A Mobile Speech, Video and Data Transceiver Scheme

L. Hanzo, R. Stedman, R. Steele, J.C.S. Cheung *

Dept. of Electr. and Comp. Science, Univ. of Southampton, U.K. S09 5NH

Abstract

The complexity, robustness, image and speech quality as well as packet multiplexing issues of a re-configurable multi-media mobile communicator are addressed. The proposed moderate complexity motion-compensated Discrete Cosine Transform (DCT) based image communicator provides an image peak signal-to-noise ratio (PSNR) of 38 dB at an average bit rate of about 25 kbits/s. The speech codec used is a low-complexity 32 kbit/s CCITT G721 standard scheme. Bandwidth efficient 16 or 64-level quadrature amplitude modulation (QAM) combined with embedded low-complexity binary Bose-Chaudhuri-Hocquenghem (BCH) forward error correction (FEC) coding is deployed. The 20-slot packet reservation multiple access (PRMA) scheme used supports an extra 2.4 kbit/s low-rate data channel for each speech user, in addition to providing 5-6 videophone channels. The ADPCM/DCT/BCH/16-QAM and ADPCM/DCT/BCH/64-QAM schemes provide nearly unimpaired speech and image quality for channel SNRs in excess of 30 dB and 38 dB, respectively.

1 Multi-media Communication

The near future will witness the integration of computation and communication in the form of highly intelligent shirt pocket sized multi-media communicators that are well endowed with computing power, memory and networking facilities, in order to serve business and, ultimately, personal users on the move. Some elements of this system, such as portable personal computers or personal mobile radio voice and data communicators are already commercially available, but have not yet shrunk to pocket calculator size and require bulky batteries.

In order to achieve the above objectives benign non-dispersive pico-cellular line-of-sight channel conditions maintaining high signal-to-noise (SNR) and signal-to-interference (SIR) must be ensured. Under these circumstances bandwidth-efficient multi-level modulation schemes can be invoked, which contribute towards supporting a high number of users generating high traffic densities. In case of low transmitted power levels the lower power efficiency of the associated linear class-A power amplifiers is less critical than that of the digital base-band signal processing.

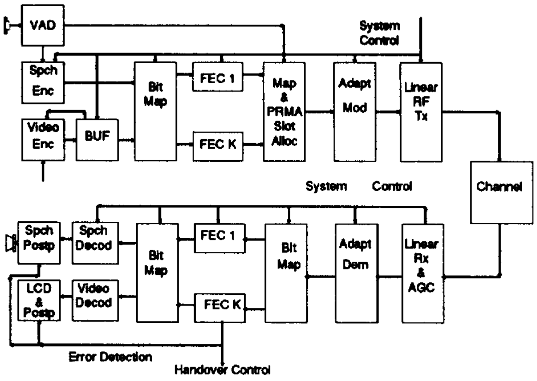

Figure 1: Multi-media Communicator Architecture

Focussing our attention on the multi-media communicator scheme displayed in Figure 1, the voice activity detector (VAD) [1] is invoked to control the packet reservation multiple access (PRMA) slot allocator [2], [1]. A further task of the 'PRMA Slot Allocator' is to multiplex digital source data from facsimile and other data terminals with the speech as well as graphics and other video signals to be transmitted. Control traffic and system information is carried by packet headers added to the composite signal by the 'Bit Mapper' before K-class source sensitivity-matched forward error correction coding (FEC) takes place. Observe that the 'Video Encoder' supplies its bits to an adaptive buffer (BUF) having a feed-back loop. If the PRMA video packet delay becomes too high or the buffer fullness exceeds a certain threshold, the video encoder is instructed to lower its bit rate, implying a concomittant dropping of the image quality.

The Bit Mapper assigns the most significant source coded bits (MSB) to the input of the strongest FEC codec, FEC K, while the least significant bits (LSB) are protected by the weakest one, FEC 1. K-class FEC coding is used after mapping the speech and video bits to their appropriate bit protection classes, which ensures source sensitivity-matched transmission. 'Adaptive Modulation' is deployed [3, 4], with the number of modulation levels, the FEC coding power and the speech/video source coding algorithm adjusted by the 'System Control' according to the dominant propagation condi-

*Currently Dept. of Electrical and Electronic Engineering, University of Bristol

0-7803-1927-3/94/$4.00 © 1994 IEEE

tions, bandwidth and power efficiency requirements, channel blocking probability or PRMA packet dropping probability. If the communications quality or the prevalent system optimisation criterion cannot be improved by adaptive transceiver re-configuration, the serving base station (BS) will hand the portable station (PS) over to another BS providing a better grade of service. Our current research is targetted at specific control algorithms.

After this rudimentary system-level introduction let us now focus our attention on the video compression and communications aspects of our proposed adaptively re-configurable communicator.

2 Video Codec

Codec Outline The outline of the proposed discrete cosine transform (DCT) based image codec is shown in Figure 2. It achieves lossy data compression by quantising and transmitting the DCT coefficients of the interframe motion compensated prediction (MCP) difference signal. A basic block matching motion compensation algorithm is used to predict a block of the source frame to be encoded by the help of the previous locally reconstructed block, which is identical to the decoded block of the decoder. The MCP error between the source block to be encoded and its prediction is discrete cosine transformed.

The Gaussian distributed DCT transform coefficients $S(m,n)$ of the MCP error signal are normalised by their standard deviations $a_{(m,n)}$ in order to be subjected to zero-mean, unit-variance Max-Lloyd scalar quantisation (SQ) using $b_{(m,n)}$ number of bits, where $b_{(m,n)}$ is dependent upon the coefficient energy distribution over the block. The DCT coefficients' energy distribution is classified into one of N energy distribution classes, whose centroids can be computed by the help of a training set using the Linde-Buzo-Gray clustering algorithm [5]. Then every DCT block to be encoded is classified into one of the N classes using the minimum mse criterion and the corresponding class-specific scaler quantiser look-up table is invoked. Finally the side information constituted by the motion vectors (MV) and block classifiers are multiplexed with the quantised MCP error signal to a single stream ready for transmission.

The quantised MCP error signal is also inverse quantised in the block SQ^{-1}, multiplied by its standard deviation and the inverse DCT is computed to yield the locally decoded MCP error signal. The MCP error signal is added to the predicted source frame in order to form the locally reconstructed frame that is to be used in the next motion compensation step. The following paragraphs describe the DCT codec in greater detail.

Source Frame Prediction The source frame is uniformly subdivided into rectangular blocks for motion compensation. A computationally efficient tree-structured suboptimum block matching motion compensation algorithm [6] is used to find the best matching block in the previous frame. Matching is based on the least mean squared error (mse) metric. The spatial offset between the source and prediction block referred to as the motion vector (MV) is encoded and transmitted as side information. For the simulations the block sise was set to 16x16 pels, with a search window of ±3 pels [6].

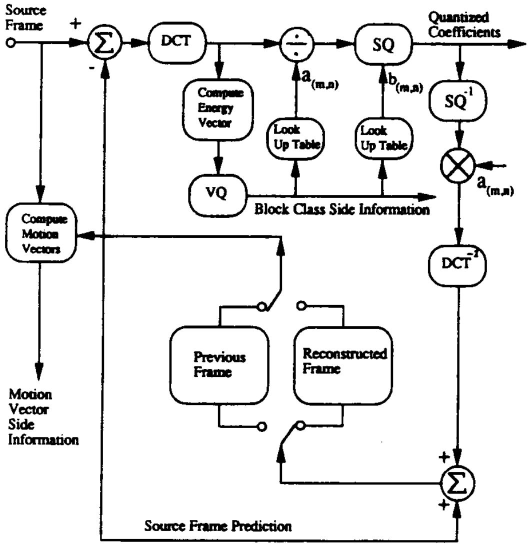

Figure 2: Schematic diagram of the interframe DCT codec

Transform Domain Quantisation and Encoding The coefficient quantisation algorithm aims to minimise the expected quantisation noise power under the constraint of a given total number of quantisation bits. In order to adaptively adjust the quantisation bit allocation scheme for each block, the coefficient energy distribution was computed over the block. Each block was then classified as one of N possible energy distribution classes and for each class an optimum quantiser was designed. The DCT coefficients of the MCP error signal are typically Gaussian distributed for all classes of blocks, but the coefficients' standard deviation varies with class. Each DCT coefficient $S(m,n)$ is quantised to $b_{(m,n)}$ number of bits accuracy by a Max-Lloyd quantiser, where $m,n = 1 \dots 16$. Rate-distortion theory for a Gaussian source states that in order to achieve a mean distortion no greater than D, the minimum number of quantisation bits required has to satisfy the following condition:

$$b(m,n) = \frac{1}{2}\log_2\left(\frac{a^2_{(m,n)}}{D}\right). \quad (1)$$

Thus given the distortion parameter D and tables of coefficient standard deviation $a_{(m,n)}$, the required quantisation precision $b_{(m,n)}$ of each DCT coefficient can be computed. The allocation of bits depends on the classification of blocks, which is based on their energy distribution.

Transform Energy Classification The block classification algorithm is based on vector quantisation of a K dimensional *energy vector* $V = (v_0, v_1, \dots, v_{K-1})$, which is derived from the DCT coefficient energy distribution. Each dimension v_l of the energy vector V is formed from the average of DCT coefficient energies over a region R_l constituted by N_l

DCT coefficients, where

$$v_l = \frac{1}{N_l} \sum_{(m,n)\in R_l} S^2(m,n). \quad (2)$$

The energy vectors computed from a training set are then used to train a codebook, which can be invoked in the block classification process. Explicitly, the blocks to be classified are compared to the energy vector centroids derived by the well known Linde, Buso, Gray clustering algorithm [5] and assigned to a particular class, closest to it in the mean squared sense. This way each DCT coefficient block characterised by its energy distribution vector is assigned to a particular energy distribution class, which is described by the help of its vector centroid. Since this classifier determines the bit allocation scheme for the block, it must also be transmitted to the decoder.

The number of energy distribution classes N was determined by means of computer simulation studies using 176 × 144 pels Quarter Common Intermediate Format (QCIF) head-and-shoulder video telephone sequences. The mean reconstructed image peak signal to noise ratio (PSNR) was computed for three different DCT coefficient masks, for $N = 2,4,8$ and 16 different energy distribution classes associated with differently trained quantisation code books and distortion values of $D = 3,6,10,25$. The PSNR varied over the range of 38-43 dB and we decided to adopt a $N = 2, D = 25$, 0.05 bits/pel scheme.

The average side information rate including the MVs and the DCT energy distribution classifier for the Miss America sequence was 930 bits/frame, while the average MCP error signal coding required 1581 bits/frame. The average total DCT codec transmission rate for a frame repetition rate of 10 frames/sec became approximately 25.1 kbps. Having described the video compression algorithm let us now concentrate on aspects of multiplexing video, speech and data sources having different statistical properties for transmission via mobile radio channels.

3 Multi-media Packet Reservation Multiple Access

Wong and Goodman studied the performance of a mixed speech and data PRMA scheme in reference [2], while the objective speech performance of a PRMA assisted cordless telecommunications (CT) speech scheme was reported in reference [1]. In contrast to previous studies, here we portray PRMA as a multimedia packet multiplexer for conveying a mixture of speech, data and video information from mobile multimedia PSs to the BS on a demand basis, exploiting the different statistical properties of bursty speech, data and video sources.

In case of speech transmissions the voice activity detector (VAD) [1] of the PS queues speech packets of the speech encoder to contend for an up-link TDMA time-slot and if the PS gets permission to transmit, it reserves its time-slot for future transmissions. When the VAD detects a silent gap, it surrenders the time-slot for other speech, data or video users, who are becoming active. If a speech user cannot get a reservation within 30 ms, its packet must be dropped but the dropping probability must be kept below $P_{drop} = 1\%$ [2].

Data users on the other hand cannot afford dropping packets at all, but tolerate longer delays and can be allocated to slots, which are not reserved by speech users in the present frame. Wong and Goodman [2] proposed an Integrated PRMA (IPRMA) protocol in which data users contend for a number of subsequent vacant slots determined from the total number of vacant slots in the frame and from the number of packets awaiting transmission in the data queuing buffer.

Similarly to speech packets, data packets are assigned a permission probability that either allows or refuses permission to contend during any particular slot for a reservation within the current TDMA frame. An interesting feature of the IPRMA scheme is that it is advantageous to allow data users to contend only, if they have a certain minimum number of packets in their buffers in order to prevent too frequent contentions and collisions for the sake of delivering a single packet. While speech stability is maintained until $P_{drop} < 1\%$, data transmission stability is defined in terms of maximum data delay or buffer length.

Using our multimedia PRMA multiplexer, in addition to 32 kbit/s adaptive differential pulse code modulated (ADPCM) speech and low-rate data users, we also accommodated a variable number of mobile video phone users, who were not allowed to drop video packets at all, in order to prevent the loss of video synchronisation. The video packets contended for reservation immediately without imposing a minimum buffer content condition, when a video packet reached the contention buffer. This allowed us to minimise the video delay and retain adequate 'lip-synchronisation' for voice and video. Further details on the multi-media PRMA multiplexer can be found in references [2, 1, 7].

In order to study the performance of our multi-media packet multiplexer, the PRMA channel rate was 720 kbps, we used 20 time slots of 1 ms duration, a 64-bit signalling header, as well as 32 kbps CCITT G721 ADPCM-coded speech. Initially no video users were served and the optimum speech permission probability was $P_{sp} = 0.3$, while the data permission probability $P_d = 0.03$ [2]. System stability was defined as $P_{drop} < 1\%$, and a maximum data buffer length of 200 packets, each delivering 640 bits. The total maximum data storage required was then 200 × 640/8=16 KBytes. At the stability limit the maximum supported data rate was found to be 13.6 kbps, when assuming 20 speech plus 20 data users. Explicitly, due to PRMA we provided an extra 13.6 kbps data channel for each user in addition to unimpaired speech communications, which was equivalent to a total channel capacity gain of $20 \times 13.6 = 272$ kbps or a relative gain of $272/640 = 42.5\%$, normalised to the primary information rate of $20 \times 32 = 640$ kbps. When defining slot occupancy as the proportion of actively exploited time slots, the slot occupancy at a speech activity of 42% became $(20 \times 0.42 \times 32 + 20 \times 13.6)/640 = 84.5\%$, which was achieved due to the fact that only the 20 speech users had a strict delay limit of 30 ms, while the remaining data users were found to experience a 700 ms average delay.

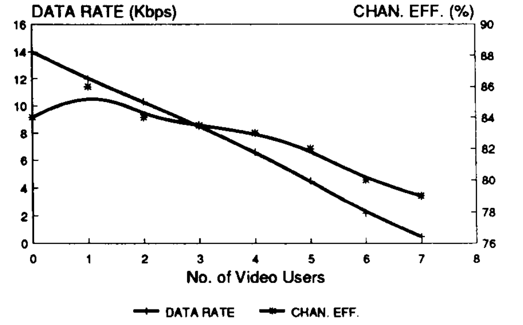

Figure 3: PRMA data rate and channel efficiency versus number of video users

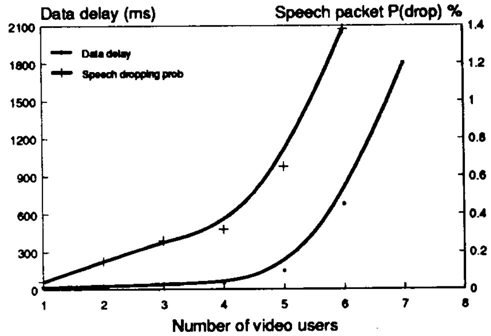

Figure 4: PRMA data delay and speech packet dropping probability versus number of video users

In our next endeavour we included a variable number of video users, whose stability was defined as an average packet delay below 100ms and no packet dropping was tolerated. The overall system performance is characterised with the aid of Figure 3, where the maximum data rate and the slot occupancy are plotted against the number of video users supported. As expected, the supported data rate was inversely proportional to the number of video channels, while the highest slot occupancy was attained with only a low number of video subscribes. This was explained by the added priority needed for the video users in order to maintain a packet delay below 100 ms, which required a higher proportion of vacant slots so that an earlier reservation became more likely. Furthermore, the data messages tolerated higher delays and hence were conveyed using a lower proportion of vacant slots. These tendencies can be confirmed using Figure 4 in terms of the delay of 2.4 kbps data users and speech packet dropping probability versus number of video users, which are scaled on the left and right hand side axes, respectively. When the number of video users exceeds five, both the speech packet dropping probability and the data delay rise sharply, infringing the stability limits for seven video suscribers. Let us now turn towards the issues of source-sensitivity matched FEC-coding, modulation and transmission over the fading mobile channel.

4 Adaptive Multi-media Link

Returning to Figure 1, the more vulnerable motion vector bits as well as the DCT block classifier bits were assigned a binary Bose-Chaudhuri-Hocquenghem (BCH) BCH(63,30,6) code, while the less sensitive motion compensated prediction error coding bits constituted by the DCT coefficients were protected by a BCH(63,51,2) code. Interleaving over an image frame was deployed in order to disperse the bursty errors caused by the Rayleigh-fading channel.

This bit stream was then transmitted using low-complexity, differentially encoded 16-QAM and 64-QAM modems [3]. The more sensitive bit class gave an average of $930 \times 63/30 = 1953$ bits/frame, and similarly, the less sensitive class yielded $1581 \times 63/51 = 1953$ bits/frame. The total average transmission rate at 10 frames/sec frame refreshing frequency became 39.06 kbps. When using 16-QAM, the single-user video signalling rate was about $39.06/4 = 9.765 \approx$ 10 kBd, while in case of 64-QAM about 6.5 kBd. Using raised cosine Nyquist filtering with an excess bandwidth of 50 %, the single-user video bandwidth requirements are about 15 and 10 kHz for the 16-QAM and 64-QAM modems, respectively.

In order to maintain low complexity here we opted for the standard CCITT G721 speech codec and the 32 kbps ADPCM-coded speech stream was FEC-coded using a BCH(63,39,4) code having a coding rate of $R = 39/63 \approx 0.62$, which yielded an FEC-coded speech rate of 51.7 kbps. The speech signal was interleaved over a 20 ms speech frame, which did not introduce any additional delay. The single-user signalling rate in case of 16-QAM became approximately 12.9 kBd while for 64-QAM 8.6 kBd.

In order to accommodate the redundancy bits due to FEC, the PRMA multi-user rate was increased according to the redundancy rate of the primary information, namely speech, yielding a PRMA rate of $1/R \times 720kbps \approx 1.61 \times 720kbps =$ 1161 kbps. When using 4 bits/symbol 16-QAM, the multi-user PRMA signalling rate becomes $1161/4 \approx 290$ kBaud, while in case of 64-QAM about 194 kBaud. Since the single-user video-rate is slightly lower than the speech-rate, the video packets are readily accommodated by the PRMA slots designed for the speech packets. The corresponding transmission bandwidth requirements in case of the above Baud rates and 50 % excess bandwidth are 435 kHz and 294 kHz, respectively. Both of these bandwidths remain substantially below the typical micro-cellular channel coherence bandwidth and hence do not require channel equalisation.

The overall image PSNR versus channel SNR performance of our multi-media video communicator is depicted in Figure 5 for both 16-QAM and 64-QAM, when using a pedestrian PS speed of 2 mph, propagation frequency of 1.9 GHz and Baud rates of 290 and 194, respectively. For 16-QAM a channel SNR in excess of about 30 dB was required to provide nearly unimpaired image quality associated with a PSNR in excess of about 36 dB. Below 30 dB channel SNR the image quality rapidly degraded. For the more bandwidth efficient but less robust 64-QAM modem the channel SNR needed to be above 38 dB for the multi-media PS to deliver unimpaired image quality. The robustness of the ADPCM speech codec scaled on the right hand side axis in Figure 5 in terms of Segmental SNR (SEGSNR) was also similar to that of the video codec, requiring a channel SNR in line with that necessitated by the video codec.

Observe in Figure 1 that the DCT codec feeds its information to a buffer (BUF) having an adaptive feedback to the codec. This buffer serves two purposes. Firstly, if the DCT codec temporarily produces too high a bit rate in its attempt to maintain the target image distortion D, the bit rate fluctuation can be smoothed by the help of the buffer. If the buffer fullness exceeds a threshold, the adaptive feedback can instruct the DCT codec to invoke a different set of quantisation look-up tables that has a higher target distortion D and hence a lower average bit rate. Secondly, also the PRMA packet multiplexer can lower the number of DCT codec bits via its feedback to the buffer, should the video delay become temporarily too high either due to the high number of users wishing to access the system or due to temporarily high DCT source rate.

Should any of the communications quality parameters evaluated in Figure 1 degrade, the adaptive multi-media transceiver can re-configure itself and reduce the number of modulation levels to 16 from 64, or even further. It is also possible to reduce the number of DCT source bits generated, in order to allocate more bits for FEC. Alternatively, hand-over to an adjacent cell ensuring better communication quality can ensue. Traffic-motivated hand-overs are also possible for example if the BS's PRMA multiplexer detects too high video packet delays or speech packet dropping probability due to collisions. However, if the communications integrity improves, the number of modulation levels can be increased again, allowing the transceiver to transmit at a higher source rate ensuring higher speech and video quality. The number of modulation levels is always communicated as side-information to the receiving end. It is feasible to use different number of bits per modulation symbol in different time slots, but the exact definition of specific adaptive system control algorithms is the subject of our current research [3].

5 Conclusion

The complexity, robustness, image quality and bit rate issues of an adaptive multi-media communicator have been addressed. If the channel SNR is in excess of about 38 dB, our 194 kBd 20-user PRMA/64-QAM/BCH/DCT communicator requires a bandwidth of about 294 kHz and provides nearly unimpaired speech and image quality associated with an image PSNR of about 38 dB and speech SEGSNR of 20 dB. If the channel SNR or other channel parameters degrade, the adaptive communicator of the near future will detect the increasing BER or SIR and drops the number of modulation levels to 16 or even further, thereby reducing the channel SNR requirement to about 30 dB or below. PRMA allowed us to support an extra 13.6 kbps data channel for each of the 20 users, or a combination of 20 speech and 2.4 kbps data subscribers, as well as 5-6 video users. Our current research is targeted at defining specific adaptive system control algorithms in order to achieve best performance amongst dynamically changing conditions and the interested reader is referred to Chapters 13, 17 and 18 of reference [3] or to [4].

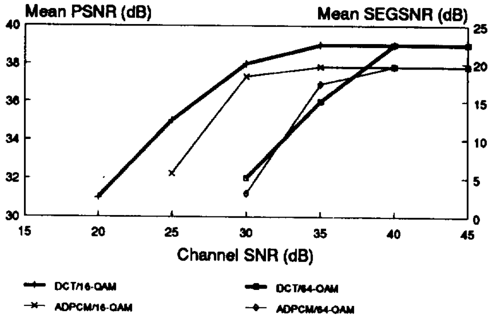

Figure 5: Image PSNR and speech SEGSNR versus channel SNR performance using 16-QAM and 64-QAM

Acknowledgement: The financial support of the SERC, UK (GR/J46845) is gratefully acknowledged.

References

[1] L. Hanzo, J.C.S Cheung, R. Steele, W.T. Webb: *A packet reservation multiple access assisted cordless telecommunications scheme,* to appear in IEEE Tr. on Veh. Techn., 1993

[2] W. Wong, D. Goodman: *A packet reservation access protocol for integrated speech and data transmission,* Proc. of the IEE, Part-I, Dec. 1992, Vol 139, No 6, pp 607-613.

[3] W.T. Webb, L. Hanzo: *Quadrature Amplitude Modulation,* Pentech Press, London, 1994

[4] R.Steele, W. T. Webb: *Variable rate QAM for data transmissions over mobile radio channels,* Keynote paper, Wireless 91, Calgary Alberta, July 1991

[5] Y. Linde, A. Buzo, R.M. Gray *An Algorithm for Vector Quantiser Design,* IEEE Trans. on Comms., Vol. Com-28, No1. Jan., 1980

[6] R. Stedman, H. Gharavi, L Hanzo, R. Steele, *Transmission of Subbband-coded Images via Mobile Channels,* IEEE Tr. on Circuits and Systems for Video Technology, Febr. 1993, Vol. 3, no.1, pp 15-27

[7] M. Eastwood, L. Hanzo, J.C.S. Cheung: *Packet reservation multiple access for wireless multimedia communications,* Electr. Lett., 24-th June, 1993, Vol. 29.,No.13, pp 1178-1179