An Adaptive Mobile Multi-media Communicator

R. Stedman, L. Hanzo, R. Steele

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

Abstract

The complexity, robustness, image quality as well as bit rate and power consumption issues of adaptive multi-media communicators are addressed. The proposed robust, 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. Bandwidth efficient adaptive 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. Our proposed 6.5-10 kBd adaptive multimedia communicator adjusts the number of modulation levels, the image and FEC codec parameters as well as other system algorithms in order to meet time-varying optimisation criteria. The DCT/BCH/16-QAM and DCT/BCH/64-QAM schemes provide nearly unimpaired image quality for channel SNRs in excess of 30 dB and 40 dB, respectively.

1 Introduction

It is anticipated that 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.

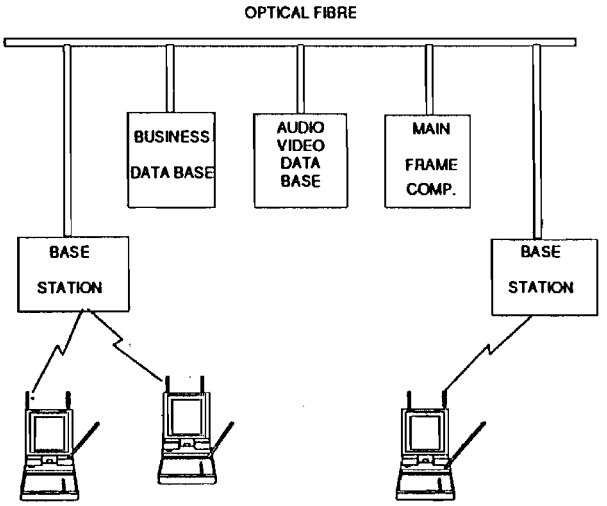

Figure 1: Integrated Multi-media Network

The multi-media personal stations (PS) seen in Figure 1 communicate with each other using both audio and video information. They can also access a variety of professional and commercial data bases via the wireless base station (BS) [1]. The advantages of accessing dynamically changing traffic data or transport schedules, frequently updated banking or other business information, stock lists, etc. while on the move are obvious. Future applications are in the area of electronic signatures or person identification by voice recognition, etc. Depending on the user's wish, computations can be carried out locally in case of network congestion or while roaming in a poorly covered communications cell. More computationally intensive tasks can be run on central

resources, as suggested by Figure 1, adding a further subjective dimension to finding the best compromise i terms of power consumption, communications convenience, system response time and so forth. In reference [a compact pen input device integrated with the flat screen display is proposed instead of the convention keyboard in order to enable user-friendly system access.

2 The Mobile Communication Channel

The adaptive PS multiplexes a variety of source information having different statistical properties onto mobi radio carriers of a microcellular system, which has different propagation properties from contemporary larg mobile radio cells [2]. The essential features are that the transmitted power is very low, typically of the order a few mWatts and there is often a dominant line of sight propagation path between the BS and PS, which result in a benign Rician received signal envelope fading, as opposed to the more violent Rayleigh fading of large ce systems. Furthermore, due to the low transmitted signal power there are no significant far-field wave reflection and hence the channel's impulse response is sharply decaying. In wave propagation parlance this is expresse as low signal dispersion, which removes the requirement of complex and power-hungry channel equalisers an robust channel codecs. This is an essential step towards power, complexity and weight reduction.

This propagation scenario is prevalent in the indoors cells of cordless telecommunications (CT) systems encoun tered in business centres, railway stations, airports, etc., where it is possible to provide relatively high signa to noise ratios (SNR). The partitioning walls provide the natural pico-cell boundaries, curtailing co-channe interference imposed upon other users communicating using the same radio frequency. The typically high SN values allow the deployment of highly bandwidth-efficient multi-level modulation schemes, which contribut towards supporting a high number of users generating high traffic densities. At these low power levels th lower power efficiency of the associated linear class-A power amplifiers is readily tolerated. Alternatively, mor efficient non-linear class-C amplifiers can be utilised combined with appropriate pre- or post-distorters [3].

Throughout our experiments in this contribution we used a wave propagation frequency of 1.9 GHz, the fr quency of the emerging personal communications networks and a pedestrian multi-media PS speed of 2 mp The signalling rate was adjusted to 1037 kBd for our 4 bits/symbol 16-QAM [4] and 6 bits/symbol 64-QAM [modems so that the transmitted speech and video multiplex signals could be accommodated in a 1728 kH wide channel slot used by the Digital European Cordless Telephone (DECT) scheme. The DECT system wa designed for office-type indoors cells similar to our multimedia system and it does not have a channel equalise since the pico-cellular multi-media environment has no significant signal dispersion at the 1152 kHz DEC signalling rate.

3 Multi-media Communicator Architecture

The pivotal implementational point of such a multi-media PS is that of finding the best compromise among a number of contradicting design factors, such as power consumption, robustness against transmission error spectral efficiency, audio/video quality and so forth. In this contribution we will address a few of these issue mainly concentrating on the video data compression and communications aspects of the proposed PS depicte in Figure 2. The time-variant optimisation criteria of a flexible multi-media system can only be met by a adaptive scheme, comprising the firmware of a suite of system components and loading that combination speech codecs, video codecs, embedded channel codecs, voice activity detector (VAD) and modems, which fulfil the prevalent one [6]. A few examples are maximising the teletraffic carried or the robustness against chann errors, while in other cases minimisation of the bandwidth occupancy, the blocking probability or the pow consumption is of prime concern.

Focussing our attention on the adaptive speech and video links displayed in Figure 2, the voice activity detect (VAD) [7] is deployed to control the packet reservation multiple access (PRMA) slot allocator [8], [6]. A furth task of the 'PRMA Slot Allocator' is to multiplex digital source data from facsimile and other data termina with the speech as well as graphics and other video signals to be transmitted. Control traffic and syste information is carried by packet headers added to the composite signal by the 'Bit Mapper' before forward err correction coding (FEC).

The Bit Mapper assigns the most significant source coded bits (MSB) to the input of the strongest FEC code FEC K, while the least significant bits (LSB) are protected by the weakest one, FEC 1. K-class FEC coding used after mapping the speech [5] and video bits [4] to their appropriate bit protection classes, which ensur source sensitivity-matched transmission. 'Adaptive Modulation' is deployed [6], with the number of modulati

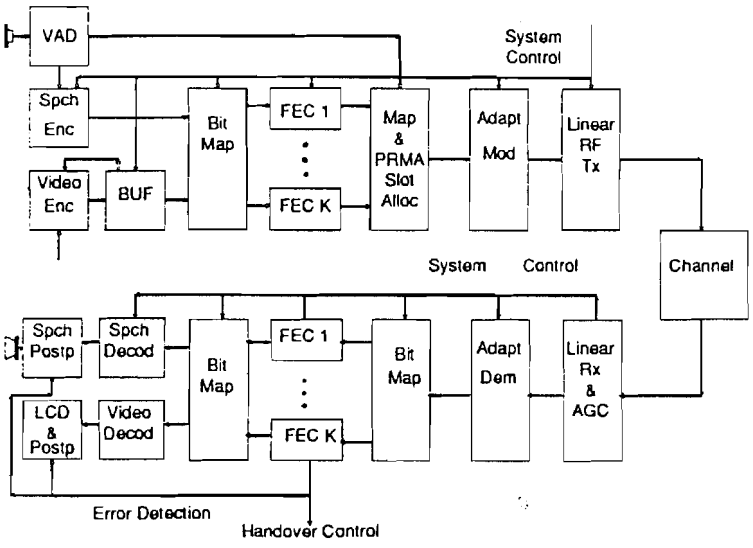

Figure 2: Multi-media Communicator Schematic

levels adjusted by the 'System Control' according to the dominant propagation conditions, bandwidth and power efficiency requirements, channel blocking probability or PRMA packet dropping probability. These decisions along with 'Handover' requests to other cells providing better communications quality can be based on the various statistical parameters, such as average bit error rate (BER), frame error rate (FER), received signal power, distance from the BS, interference level from co-channel users in adjacent cells, etc. that are evaluated and monitored by both the BS and PS. One of the most important parameters used to control these algorithms is the error detection flag of the FEC decoder of the most significant bit (MSB) class of speech and video bits. This flag can also be invoked to control the speech and video 'Postprocessing' algorithms. The adaptive modulator transmits the user bursts from the PS to the BS using the specific PRMA slot allocated by the BS for the PSs speech, data or video information via the linear radio frequency (RF) transmitter (Tx).

The receiver structure essentially follows that of the transmitter. After linear class-A amplification and automatic gain control (AGC) the 'System Control' information characterising the type of modulation and the number of modulation levels must be extracted from the received signal, before demodulation can take place. This information also controls the various internal bit mapping algorithms and invokes the appropriate speech and video decoding as well as FEC decoding procedures. After 'Adaptive Demodulation' at the BS the source bits are mapped back to their original bit protection classes and FEC decoded. As mentioned, the error detection flag of the strongest FEC decoder, FEC K, is used to control handovers or speech and video postprocessing. The FEC decoded speech and video bits are finally source decoded and the recovered speech arrives at the earpiece, while the video information is displayed on a flat liquid crystal display (LCD).

The system control algorithms of the adaptive mobile multimedia communicator will dynamically evolve over the years. PSs of widely varying complexity will coexist, with newer ones providing backward compatibility with existing ones, while offering more intelligent new services and more convenient features. After this system-level introduction let us now focus our attention on the video data compression and communication aspects of our proposed multimedia communicator.

4 Adaptive Video Link

4.1 Interframe DCT Video Codec

Codec Outline The construction of our adaptive video link is based on Figure 2. The attributes of the video codec strongly influence the properties of the entire multimedia system. Our video codec is designed for fading mobile channels and has high innate robustness against channel errors at a low bit rate and moderate algorithmic complexity. Its outline is shown in Figure 3. It achieves lossy data compression by quantising and transmitting the discrete cosine transform (DCT) coefficients of the interframe motion compensated prediction

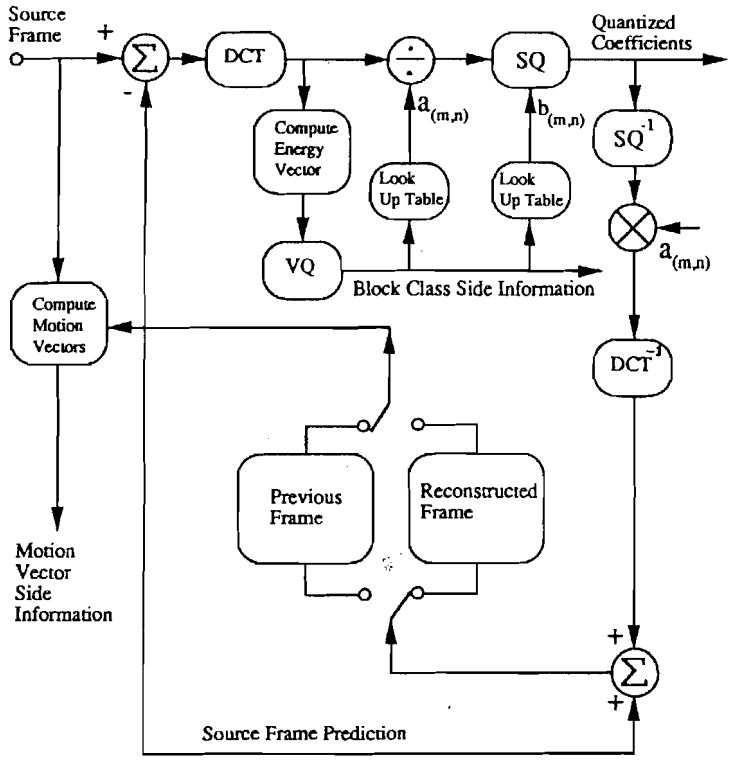

Figure 3: Schematic diagram of the interframe DCT codec

(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. 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 [4] 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 size was set to 16x16 pels, with a search window of ± 3 pels [4].

MCP Error Signal Characteristics The motion compensated frame difference (MCFD) signal's characteristics are not stationary. Areas of the image where linear translation is a good approximation to motion produce MCFD blocks with low signal energy. These blocks are sparsely populated with low valued errors of zero mean and low spatial correlation. In contrast, blocks which are not well matched produce error signals

with higher energy and spatial correlation. Only a fraction of the MCFD signal energy is contained in the more uncorrelated non-stationary blocks, thus the signal is suitable for encoding using the DCT.

Transform Domain Quantisation and Encoding The coefficient quantisation algorithm follows the principles proposed by Chen and Smith [9], which 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, a measure of the coefficient energy distribution is used to classify the block as one of N possible energy distribution classes. The DCT coefficients of the MCP error signal are 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. 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:

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

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* $V = (v_0, v_1, \ldots, v_{K-1})$, 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

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

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, Buzo, Gray clustering algorithm 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 174 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 training set images using codebook sizes or number of energy distribution classes of $N = 2, 4, 8$ and 16. The bottom part of Figure 4 shows the image PSNR against coding rate for the three different energy vectors depicted at the top of Figure 4. We used for all three DCT quantisation schemes a different $K = 4$-dimensional vector $V = (v_0, v_1, v_1, v_3)$. In case of Vector 1 all areas $R_l, l = 0 \ldots 3$ were of size $N_l = 1$ coefficient. For Vector 2 the region R_0 was of size $N_0 = 1$, regions $R_{1,2}$ were of size $N_{1,2} = 2$, while R_3 was of size $N_3 = 4$. Lastly, for Vector 3 the region R_0 was of size $N_0 = 4$, for sectors $R_{1,2}$ $N_{1,2} = 6$ pels were included, while in case of R_3 $N_3 = 9$ pels were used. For all three vectors a separately trained codebook was designed and deployed throughout our investigations.
Our simulation results presented in Figure 4 for the three DCT coefficient masking schemes specified by Vectors $1-3$ showed little performance difference between codebook sizes of $N = 4, 8$ and 16, whilst the performance of codebook size 2 degraded at higher rates. However, for lower coding rates, below 0.05 bits/pixel, the codebook size of $N = 2$ has similar image peak SNR performance to the scenarios using $N = 4, 8$ or 16 energy distribution classes. Observe that each curve has four points, corresponding to distortion values of $D = 3, 6, 10$ and 25 that are associated with different coding rate values in case of the four different codebook sizes.
Note also from Figure 4 that the different DCT coefficient masks represented by the energy Vectors $1-3$ had only a minor influence on the image quality measured in terms of PSNR. Furthermore, for an almost an order of magnitude distortion increment from $D = 3$ to $D = 25$ the PSNR reduction was quite modest for all three DCT masks, which was particularly true for $N = 2$ energy distribution classes at a concomitant tenfold reduction of the coding rate. This fact suggested that most of the information was carried by the motion vectors and

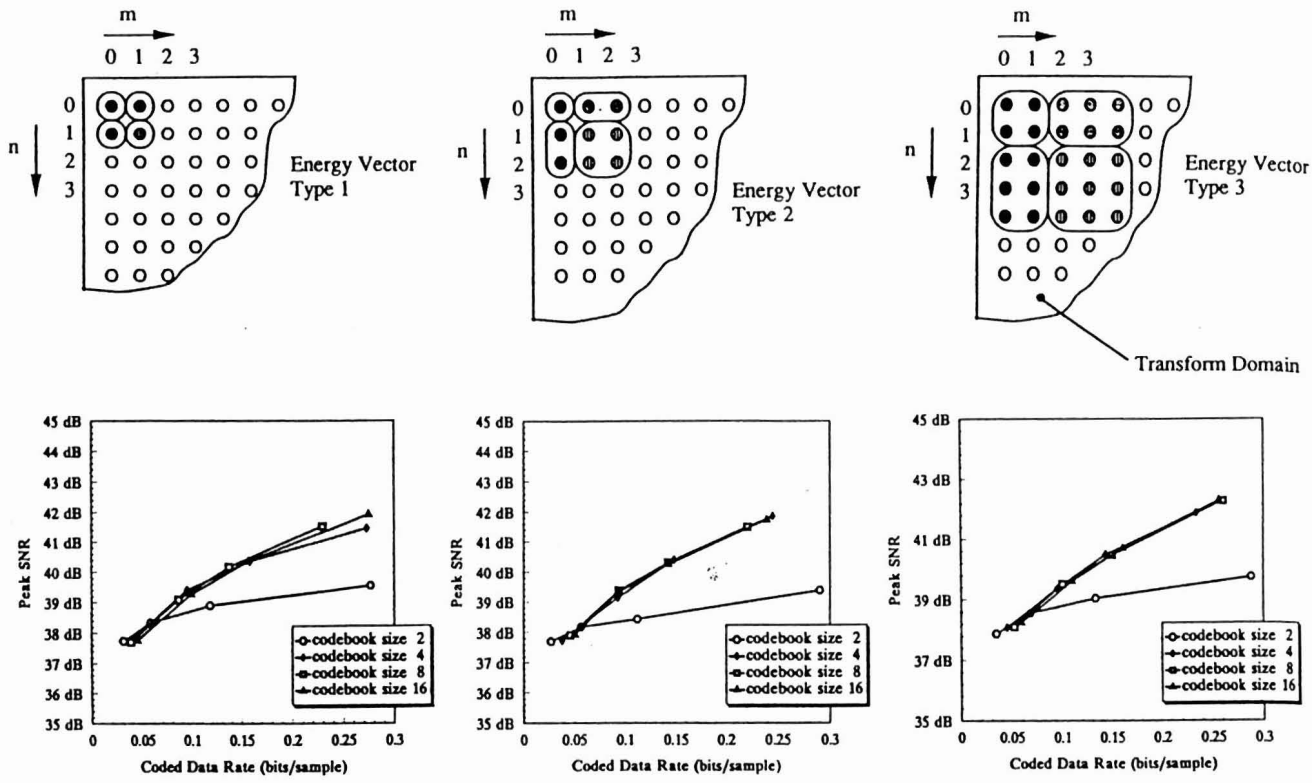

Figure 4: Three different DCT coefficient energy distribution vectors and the associated PSNR as a function of coding data rate using Vectors $1-3$

the quantisation accuracy of the MCP error signal was not particularly crucial, which allowed us to achieve a coding rate of around 0.05 bits/pel, while maintaining a PSNR of 38 dB.

In order to ensure adequate encoding of occasional non-stationary blocks having relatively wide DCT coefficient spread over a large area, finally we decided to use the DCT coefficient mask associated with Vector 3 along with its minimum MSE-optimised codebook. Figure 5 shows the fluctuation of PSNR versus image frame index, when coding the well known Miss America sequence for distortion parameter values of $D = 3, 6, 10$ and 25. Observe for example that for $D = 10$ an average PSNR of about 40 dB can be maintained at a coding rate of approximately 0.05 bits/pixel. The average side information rate including the MVs and the DCT coefficient mask 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.

4.2 Adaptive DCT/QAM Video Links

In accordance with the system block diagram seen in Figure 2 we protected the bit stream of our DCT video codec using a twin-class bit-sensitivity matched BCH FEC arrangement. The more vulnerable MV bits as well as the DCT block classifier bits were assigned a BCH(63,30,6) code, while the less sensitive MCP 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 16-QAM and 64-QAM modems. 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 became 39.06 kbps. When using 16-QAM, the signalling rate was about $39.06/4 = 9.765 \approx 10$ kBd, while in case of 64-QAM about 6.5 kBd.

The overall image PSNR versus channel SNR performance of our multi-media video communicator is depicted in Figure 6 for both 16-QAM and 64-QAM. For 16-QAM a channel SNR in excess of about 30 dB was required to provide nearly unimpaired image quality. Below this SNR value the image quality rapidly degraded. For the more bandwidth efficient but less robust 64-QAM modem the channel SNR needed to be above 40 dB for the multi-media PS to deliver unimpaired image quality.

When the DCT image codec lost synchronisation due to channel errors, this condition was detected by the

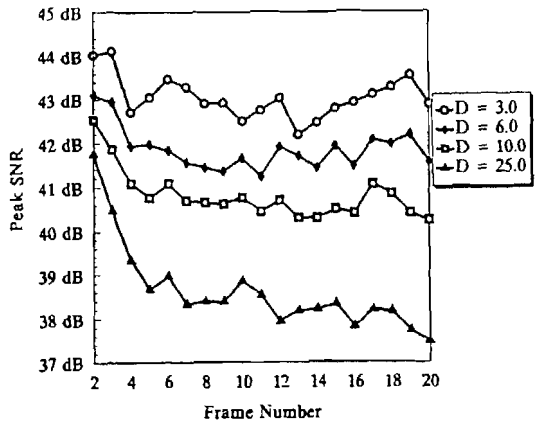

Figure 5: Image PSNR versus image frame index for the Miss America sequence for $D = 3, 6, 10$ and 25, when using the DCT mask and codebook associated with Vector 3 and $N = 4$

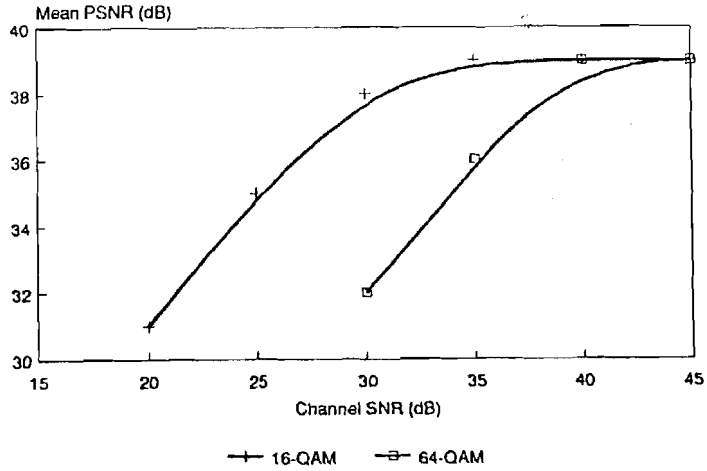

Figure 6: Image PSNR versus channel SNR for the Miss America sequence using 16-QAM and 64-QAM

BCH(63,30,6) FEC codec and bit-synchroneous operation was regained by switching to correlation mode. In this search mode the receiver continuously correlated the received signal with its locally stored 16-bit preamble and when a sharp correlation peak was detected, signalling the beginning of a new frame, synchroneous decoding resumed with the loss of only one image frame.

Observe in Figure 2 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 high generated traffic or high DCT source rate.

Should the channel conditions worsen, which can be detected by the BCH(63,30,6) codec, the adaptive multimedia transceiver can 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, handover to an adjacent cell with a higher received SNR can ensue. Traffic-motivated hand-overs are also possible for example if the BS's PRMA multiplexer detects too high video packet delays due to collisions. However, if the SNR improves, the number of modulation levels can be increased again, allowing to create more PRMA time slots and reducing the channel blocking probability. The exact definition of specific adaptive system control algorithms requires further research, although a few examples can be found in reference [6].

5 Conclusion

The complexity, robustness, image quality as well as bit rate and power consumption issues of an adaptive multi-media communicator have been addressed. The proposed adaptive multi-mode PS can communicate over a wide range of operating conditions. If the channel SNR is in excess of about 40 dB, our 6.5 kBd communicator provides nearly unimpaired image quality associated with a PSNR of about 38 dB. If the channel SNR decays, our adaptive communicator detects the increasing BER and drops the number of modulation levels to 16 or even further, thereby reducing the channel SNR requirement to about 30 dB or below. Typically slightly lower channel SNR values are necessitated by speech codecs due to their higher inherent robustness. Further research has to be targeted at defining specific adaptive system control algorithms in order to achieve best performance amongst dynamically changing conditions.

References

[1] S. Sheng, A. Chandrakashan, R.W. Brodersen, *A Portable Multimedia Terminal*, IEEE Comms. Mag., Dec 1992, Vol. 30. No. 12, pp 64-75

[2] R. Steele (Ed.), *Mobile Radio Communications*, Pentech Press, 1992

[3] S. Stapleton, F. C. Costescu, *An Adaptive Predistorter for a Power Amplifier Based on Adjacent Channel Emissions*, IEEE Tr. on Vehicular Technology, Vol. 41, No. 1. Febr. 1992, pp 49-56

[4] 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

[5] L. Hanzo, R.A. Salami, R. Steele, P.M. Fortune, *Transmission of digitally encoded speech at 1.2 kBd for PCN*, IEE Proceedings Part I, vol. 139, No. 4, August, 1992, pp. 437–447.

[6] J. Williams, L. Hanzo, R. Steele, J.C.S Cheung, *On the performance of adaptive speech terminals for UMTS*, submitted to IEEE Tr. on VT 1993.

[7] E. Bacs, L. Hanzo, *A simple real-time adaptive speech detector for SCPC systems*, Proceedings of ICC'85, Chicago, Illinois, USA, pp. 1208–1212.

[8] D.J. Goodman, S.X. Wei, *Efficiency of packet reservation multiple access* IEEE Transactions on vehicular technology, Vol. 40, No. 1, Febr. 1991, pp. 170-176.

[9] W.H. Chen, C.H. Smith, *Adaptive Coding of Monochrome and Color Images.* IEEE Transactions on Communications. vol COM-25, No 11, November 1977.