Adaptive Low-Rate Wireless Videophone Schemes

Lajos Hanzo, *Senior Member, IEEE*, and Jürgen Streit

***Abstract*— The design and performance of a range of wireless videophone transceivers are presented. Highly bandwidth efficient, fixed but arbitrarily programmable rate, perceptually weighted Discrete Cosine Transform (DCT) based video codecs are proposed for quarter common intermediate format (QCIF) videophone sequences. Perceptually weighted cost/gain controlled motion compensation and quad-class DCT-based compression is applied with and without run-length coding. Specifically, we propose video codecs having transmission rates in the range of 5–11.36 kbps and preselected the 11.36 kbps Codec 1, the 8.52 kbps Codec 2 and the 8 kbps Codec 2a, for which we designed the intelligent reconfigurable Systems 1–5. 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. Throughout these systems a partial forced update (PFU) technique was invoked in order to keep transmitter and receiver aligned amongst hostile channel conditions. 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. The system's robustness was increased using Automatic Repeat Requests (ARQ), inevitably reducing the number of users supported, which was between 6 and 19 for the various systems. In a bandwidth of 200 kHz, similarly to the Pan-European GSM mobile radio system's speech channel, using System 1 for example 16 and 8 videophone users can be supported in the 16QAM and 4QAM modes, respectively. All system features are summarized in Table III.**

I. Introduction and Motivation

IN RECENT YEARS, there has been an increased research activity in the field of videophony [33], [2]–[4], [6]–[9], [27], in particular for mobile channels. The motion pictures expert group [10], [11] (MPEG) and the p × 64 kbps CCITT H261 video codecs have been contrived for high-rate, low bit error rate fixed channels. Although the new MPEG 4 working group's activities target mobile videophony [12], to date there are no appropriate video standards for mobile videophony over existing standard radio systems.

Some authors have investigated the deployment of the H261 codec over mobile channels [5], [7] while others [2], [6], [33] have contrived a range of different run-length coded variable rate arrangements, such as subband coded schemes [9], [14]–[16], fractal-based [8] or discrete cosine transformed [27] systems. A common feature of most previously proposed systems is that they typically generate a variable output rate.

Manuscript received January 24, 1995; revised June 5, 1995. This work was supported by EPSRC, U.K., Contract GR/J46845. This paper was recommended by Associate Editor L. Wu.

The authors are with the Department of Electronics and Computer Science, University of Southampton, SO17 1BJ, U.K.

IEEE Log Number 9413981.

In contrast to the above variable rate schemes, 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 [59], the Japanese PDC [19], and the American IS-54 [20] as well as IS-95 systems [21]. The corresponding speech rates are 13, 6.7, 8, and 9.6 kbps and the proposed video codecs are capable of operating at a scanning rate of 10 frames/s, while maintaining such low bit rates. Their further advantage is that they are capable of programmable multirate operation in the 3rd generation adaptive multimedia, multimode terminal of the near future, which is currently under intensive study within the European Community's Fourth Framework Programme in the field of Advanced Communications Technologies and Services (ACTS) [13].

In this short treatise we cannot consider the performance of the proposed video codecs in all of the above 2nd and 3rd generation mobile radio systems. Our main goal is to describe the design philosophy of our prototype video codecs and document their performance using three different characteristic fixed bit rates within the above mentioned typical speech coding rate range. Furthermore, we will devote special attention to transmission robustness issues and devise and evalute a range of error control measures. In particular, a variety of unequal protection coded modulation [26], [25] and Automatic Repeat Request (ARQ) techniques are invoked and assessed in robustness, complexity, and bandwidth efficiency terms and the benefits of a particular multimode scenario are analyzed. The aspects of devising reliable transceiver reconfiguration protocols and analysing the associated switching transients as well as signalling issues are left for future studies.

Specifically, the video source rate was fixed to 8, 8.52, or 11.36 kbps, and this stream was then transmitted using an intelligent transceiver, which can configure itself as a robust but less bandwidth efficient scheme or can double its bandwidth efficiency at the cost of requiring better channel conditions. Note that although the 11.36 kbps video rate differs from the 13 kbps GSM speech rate, some of the systems proposed accommodate exactly 16 or 8 videophone users, as the full- and half-rate GSM speech system.

System Concept: According to Shannon's pioneering work [22], which was further exposed in [23] and [24], in case of lossless coding the lowest achievable source coded rate is given by the source entropy [28]. Such an ideal source encoder produces a completely uncorrelated sequence, where all symbols are mutually independent and have the same significance or error sensitivity. Any further rate reduction implies that some distortion is inflicted. Since our source codecs operate well below the source entropy, the design philosophy hinges around the principle as to how best the total distortion is distributed over the source message in the time- or frequency-domain in order to minimise its subjective effects.

When using Shannon's ideal source codecs and channel codecs over memoryless AWGN channels, where bit errors occur randomly, there is no advantage in treating source and channel coding jointly. Our nonideal source codecs however produce sequences, which still retain correlation and unequal error sensitivity. Over fading mobile channels this problem is aggravated by the bursty error statistics, which can only be randomized using infinite memory channel interleavers inflicting infinite delays. In this situation source-matched channel coding [9], [18], [23], which takes account of the source significance information [23] (SSI) brings substantial advantages in terms of reducing the required minimum channel SNR.

Joint coding and modulation in the form of trellis coded modulation (TCM) or block coded modulation (BCM) was also proposed in the literature in order to reduce the required channel SNR [51], [52], while in [18] and [25] source-matched joint source/channel coding and modulation was introduced. In this treatise we will follow a similar design philosophy in order to achieve best videophone performance over fading channels.

The schematic of the proposed transceiver is portrayed in Fig. 7 and this treatise follows the same structure. Speech source coding issues are not considered here, the reader is referred to [34] and [33] for the choice of the appropriate speech codec. Channel coding issues are addressed in [50], while a detailed discussion of modulation is given in [26]. Section II outlines the design of a variety of programmable, but fixed-rate video source codecs and analyzes their bit sensitivity. Section III details modulation and transmission aspects, which is followed by the description of the source-matched transceiver in Section IV. The system's performance is characterized in Section V, before offering some conclusions in Section VI.

II. Videophone Codecs

2.1. Codec 1

Let us initially focus our attention on the proposed discrete cosine transform [28] (DCT) based video codec depicted in Fig. 1, which was designed for hostile mobile channels. The codec uses 176×144 pixels Quarter Common Intermediate Format (QCIF) images scanned at 10 frames/s. For the sake of communications convenience and simple networking our aim was to develop a fixed-rate codec which is able to dispense with an adaptive feed-back-driven bit-rate control buffer. Therefore a constant bit-rate source codec was required, which in Codec 1 forced us to avoid using efficient variable-rate compaction algorithms, such as Huffman coding. This was achieved by fixing both the number of 8×8 blocks to be motion-compensated and those to be subjected to DCT to 30 out of $22 \times 18 = 396$. The selection of these blocks is based on a gain-controlled approach, which will be highlighted next.

In order to curtail error propagation across image frames

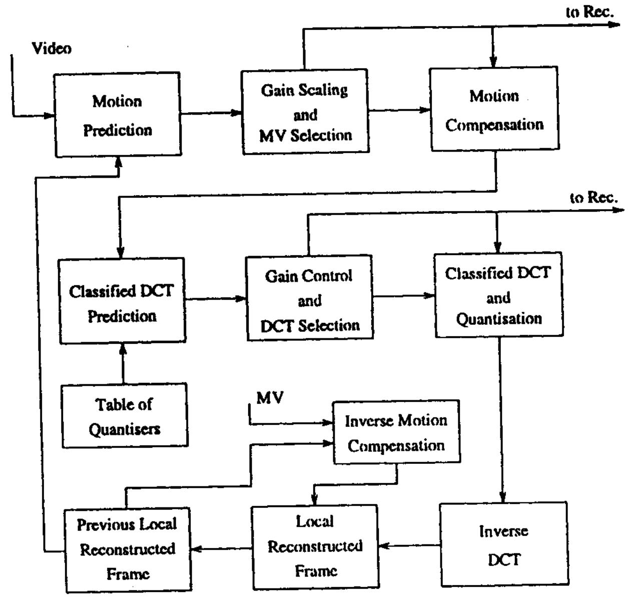

Fig. 1. Video encoder schematic.

interframe modes of operation. In the **intraframe mode** the encoder transmits the coarsely quantized block averages for the current frame, which provides a low-resolution initial frame required for the operation of the interframe codec at both the commencement and during later stages of communications in order to prevent encoder/decoder misalignment. The **interframe mode** of operation is based on a combination of gain-controlled motion compensation and gain-controlled DCT coding as seen in Fig. 1.

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 center of each block. Before the actual motion compensation takes place the codec tentatively determines the potential benefit of the compensation in terms of motion compensated error energy reduction. In order to emphasize the subjectively, more important eye and mouth region of the videophone images the potential gains for each motion compensated block are augmented by a factor of two in the center of the screen. Then the codec selects the thirty blocks resulting in the highest scaled gain, and motion compensation is applied only to these blocks, whereas for all other so-called passive blocks the codec applies simple frame differencing.

Gain Controlled Quadruple-Class DCT: Pursuing a similar approach, gain control is also applied to the DCT-based compression. Every block is DCT transformed and quantized. Because of the nonstationary nature of the motion compensated error residual (MCER) the energy distribution characteristics of the DCT coefficients vary. Therefore four different sets of DCT quantizers are available, as examplified in Fig. 2. All four bit allocation schemes are tentatively invoked in order to select the best set of quantizers resulting in the highest energy compaction gain. Ten bits are allocated for each quantizer, each of which are trained Max-Lloyd

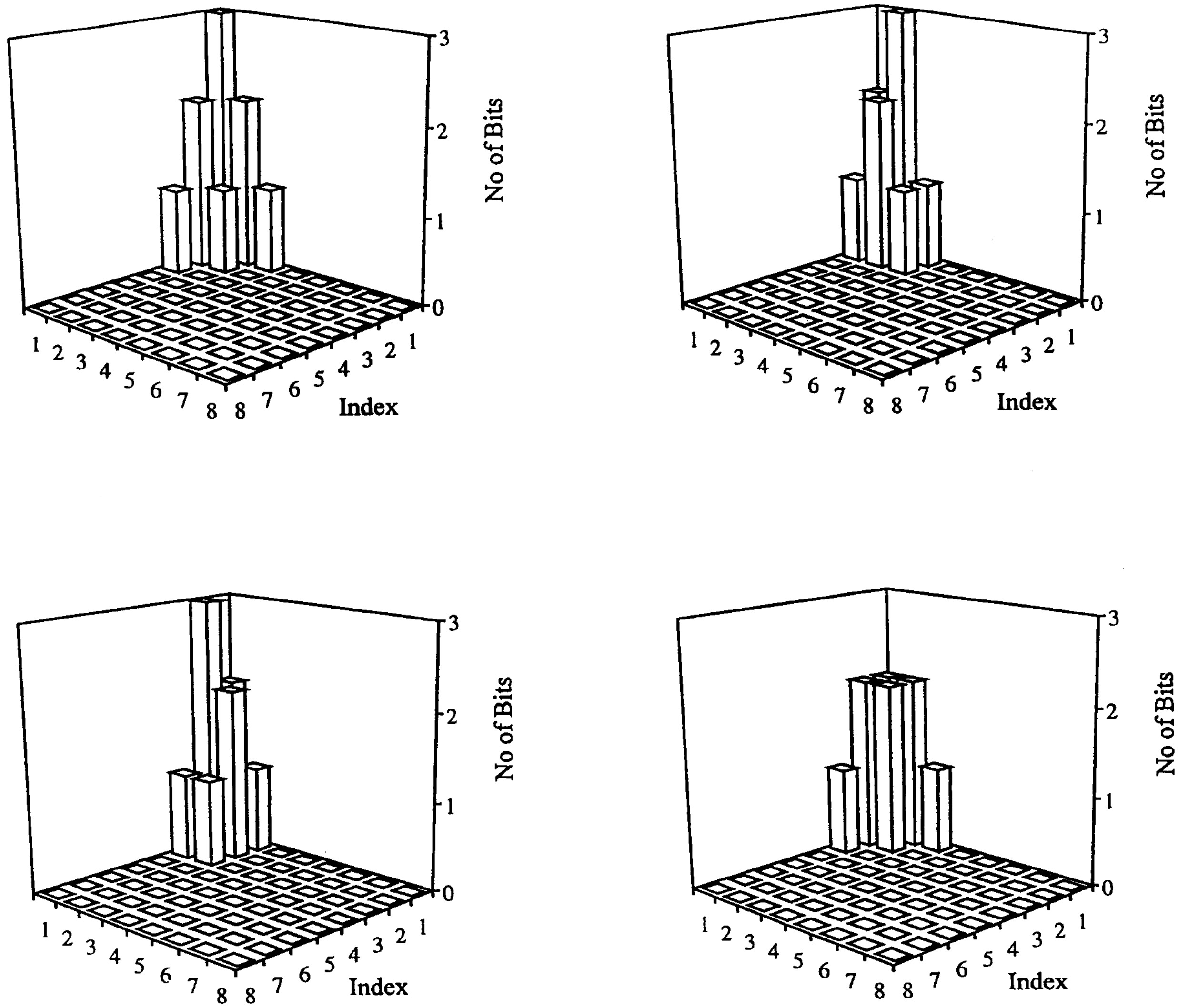

Fig. 2. Quad-class DCT quantization schemes.

distribution class. Again, the energy compaction gain values are scaled to emphasise the eye and mouth region of the image and the DCT coefficients of the thirty highest-compression blocks are transmitted to the decoder.

Partial Forced Update: The disadvantage of interframe codecs is their vulnerability to channel errors. Every channel error results in a misalignment between the reconstructed frame buffer of the encoder and decoder. The errors accumulate and do not decay, unless a leakage-factor or a partial forced update (PFU) technique is employed. In our proposed codec in every frame 22 out of the 396 blocks, scattered over the entire frame, are periodically updated using the 4-b quantized 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 de-

22-b frame alignment word (FAW). This is necessary to assist the video decoder's operation in order resume synchronous operation after loss of frame synchronization over hostile fading channels. The partial intraframe update refreshes only 22 out of 396 blocks every frame. Therefore every 18 frames or 1.8 s the update refreshes the same blocks. This periodicity is signalled to the decoder by transmitting the inverted FAW. A MV is stored using 13 b, where 9 b are required to identify one of the 396 the block indexes using the enumerative method and 4 b for encoding the 16 possible combinations of the X and Y displacements. The 8×8 DCT-compressed blocks use a total of 21 b, again 9 for the block index, 10 for the DCT coefficient quantizers, and 2 b to indicate which of the four quantizer 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 b 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

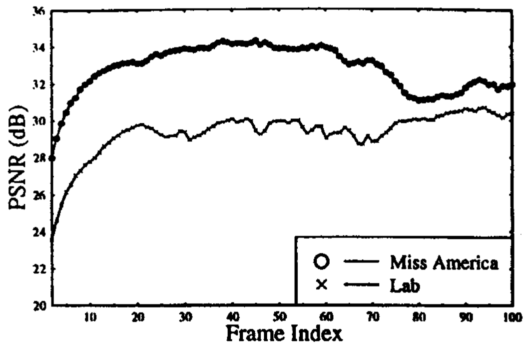

Fig. 3. PSNR performance of the 11.36 kbps Codec 1.

TABLE I
BIT ALLOCATION TABLE

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

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. The bit allocation scheme is summarized in Table I and the complexity of this codec is about 50 Mflops, which can be reduced to about 25 Mflops without significant performance penalty. In our further discourse we will refer to the above scheme as Codec 1. After addressing the bit sensitivity issues of Codec 1 we will propose a lower bit rate but more error sensitive arrangement, Codec 2, and analyze their advantages and disadvantages.

Source Sensitivity: In order to apply source-sensitivity matched protection the video bits were subjected to sensitivity analysis. In [9] we have consistently corrupted a single bit of a video coded frame and observed the image peak signal-to-noise ratio (PSNR) degradation inflicted. Repeating this method for all bits of a frame provided the required sensitivity figures and on this basis bits having different sensitivities can be assigned matching FEC codes. This technique, however, does not take adequate account of the phenomenon of error propagation across image frame boundaries. Therefore in this treatise we propose to use the method suggested in [17], where we corrupted each bit of the same type in the current frame and observed the PSNR degradation for the consecutive frames due to the error event in the current frame. As an example, Fig. 4 depicts the PSNR degradation profile in case of corrupting all 'No 1' Bits, the most significant bit (MSB) of the PFU and all 'No 11' Bits, one of the address bits of the MV, in frame 21. In the first case, the MSB of all PFU blocks are corrupted causing a scattered pattern of artifacts across the image. Those blocks will be replenished by the PFU exactly every 18 frames, revealed in the 'staircase' effect in Fig. 4. The impact of the corrupted MV is randomly distributed across the frame and hence, mitigated continuously by the PFU.

In order to quantify the overall sensitivity of any specific bit we have integrated (summed) the PSNR degradations over the consecutive frames, where they have had a measurable

[1] The MA sequence encoded at 11.36 kbps can be viewed under the address

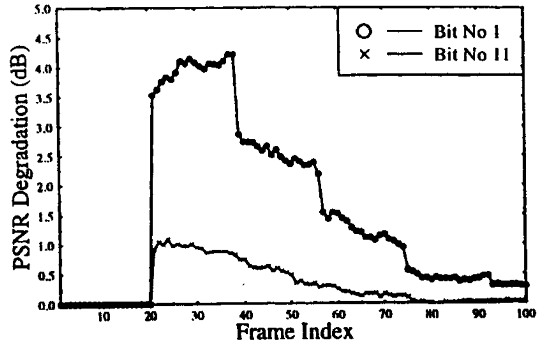

Fig. 4. PSNR degradation profile for Bits 2 and 11 of the MV in Codec 1.

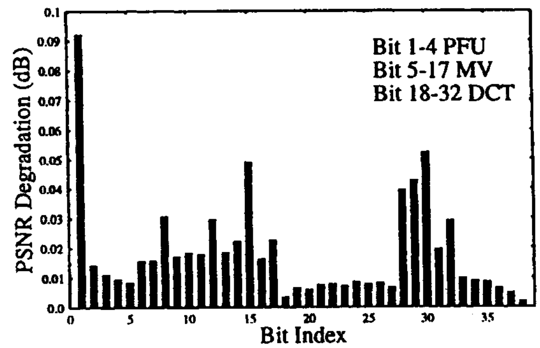

Fig. 5. Integrated PSNR bit sensitivities of Codec 1.

effect and averaged these values for all the occurrences of the corresponding bit errors. These results are shown in Fig. 5 for the 13 MV bits and 21 DCT bits of an 8 × 8 block, as well as for the 4 partial forced update bits.

2.2. *Codec 2*

In an attempt to improve the bandwidth efficiency of Codec 1 and to explore the range of design trade-offs, we have studied the statistical properties of the various parameters of Codec 1 in order to identify any persisting residual redundancy. 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.

The schematic diagram of Codec 2 is akin to that of Codec 1 shown in Fig. 7, but the above mentioned coding tables are further compressed by RLC. Similarly to Codec 1, the operation of Codec 2 is also initialized in the **intraframe mode,** where the encoder transmits the coarsely quantized block averages for the current frame. This provides a low-resolution initial frame required for the operation of the motion compensated **interframe** codec at both the commencement and during later stages of communications in order to prevent encoder/decoder misalignment. However, for the sake of maintaining a total bit rate R in the range of 5–10 kb/s for our 176×144 pixel CCITT standard QCIF images at a scanning rate of 10 frames/s we now limited the number of encoded bits per frame in Codec transmit all block averages with a 4-b resolution, as in Codec 1, while not exceeding the above stipulated maximum bit rate, we fixed the initial intraframe block size to 10×10, 12×12, or 14×14 pixels for the above three target bit rates. The intraframe block size in Codec 1 was 10×10 pixels.

However, in the motion-compensation (MC) we retained the block-size of 8×8 and the search window size of 4×4 around the center of each block. Furthermore, the previously proposed gain-controlled MC and quad-class DCT quantization was invoked. 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 packetized to a symbol and then run length encoded. As a result, a typical 396-b active/passive table containing 30 active flags can be compressed to less than 150 b. 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 MV's 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 Fig. 6 and the average results are summarized is Table II. Although due to the low-resolution intraframe mode at the commencement of communications it takes a few frames for the image to reproduce fine details, this effect is not objectionable. This is because the subjectively more important center of the screen is processed first. Fig. 6 demonstrates that at 5 kbps the codec operates at its limits and hence it takes a long time before the steady-state PSNR value is reached. However, at rates at or above 8 kbps a pleasant quality is maintained leading to an average PSNR in excess of 30 dB, which is exceeded in the center of the image. Based on these findings, in the run length coded System 2 we have opted for an **8.52 kbps** implementation of **Codec 2**, generating 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.** Before we continue with the description of the source-matched transceiver schemes it must be emphasized that, in contrast to Codec 1 where no RLC is employed, if the RL-coded activity table bits are corrupted, the rest of that frame will be completely corrupted. Hence automatic repeat request (ARQ) techniques are preferred in the systems employing the RL-coded Codec 2. The sensitivity of the remaining bits is similar to that of the corresponding Class Two bits of Codec 1.

TABLE II
AVERAGE PSNR PERFORMANCE OF CODEC 2 FOR THE 'MISS AMERICA' AND 'LAB' SEQUENCES

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

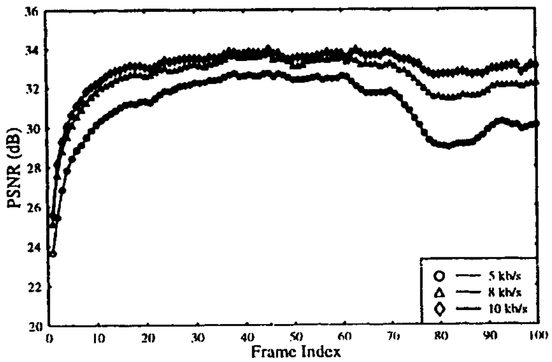

Fig. 6. PSNR versus frame index performance of Codec 2 for the 'Miss America' sequence.

III. MODULATION AND TRANSMISSION

Over mobile channels constant envelope modulation techniques, such as for example Gaussian Minimum Shift Keying (GMSK) used in the Pan-European GSM system [59] has successfully been applied. In contrast, until quite recently QAM research was mainly focused at applications over AWGN channels [35]. However, fuelled by the drive towards ever higher bandwidth efficiency and facilitated by advances, such as noncoherent star QAM [45], coherent pilot symbol assisted modulation [55] and the transparent tone in band [56], [57] (TTIB) technique, during the last few years its employment has also become realistic over mobile channels [36]–[48]. In order to achieve high bandwidth efficiency, QAM encodes information on both the phase and magnitude of the complex transmitted signal and hence it requires a linear transceiver, which suffer from low power efficiency [53], [54]. However, in low-power pico- or microcellular applications this is not a serious limitation, since the power consumption of the high-complexity digital circuitry is more crucial. In fact, due to its reduced signalling rate such a transceiver may be able to operate in a nondispersive scenario, without a channel

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS

Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS

Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS

Sync your system to PACER to automate legal marketing.