A WIDEBAND TRANSFORM-CODED SYMBOL-BY-SYMBOL ADAPTIVE SPEECH AND AUDIO TRANSCEIVER

T. Keller, M. Muenster, L. Hanzo

Dept. of Electronics and Computer Science,
University of Southampton, SO17 1BJ, UK.
Tel: +44-1703-593 125, Fax: +44-1703-594 508
Email:lh@ecs.soton.ac.uk
http://www-mobile.ecs.soton.ac.uk

ABSTRACT

The design trade-offs of burst-by-burst adaptive Orthogonal Frequency Division Multiplex (OFDM) wideband speech transceivers are analysed. A constant throughput adaptive OFDM transceiver was designed and benchmarked against a time-variant rate scheme. The proposed joint adaptation of source- and channel-coder as well as modulation regime results in attractive, robust, high-quality audio systems.

1. INTRODUCTION

In this contribution we introduce a duplex high quality, 7-KHz bandwidth audio communications system, which will be used to highlight the system design aspects of burst-by-burst adaptative modulation, channel coding and source coding.

Burst-by-burst adaptive transceivers have recently generated substantial research interests in the wireless communications community [1]-[4]. The transceiver reconfigures itself on a burst-by-burst basis, depending on the instantaneous perceived wireless channel capacity. More explicitly, the associated channel quality of the next transmission burst is estimated and the specific modulation mode, which is expected to achieve the required performance target is then selected for the transmission of the current burst. In other words, modulation schemes of different robustness and of different data throughput are invoked. For data transmission systems, which do not necessarily require low transmission delays, variable–throughput adaptive schemes can be devised, which operate efficiently in conjunction with powerful error correction codecs, such as long block length turbo codes. By contrast, fixed rate burst-by-burst adaptive systems, which sacrifice a guaranteed BER performance for the sake of maintaining a fixed data throughput, are more amenable to employment interactive speech and video communications systems.

In this paper, we propose a hybrid adaptive transceiver scheme, based on a multi–mode constant throughput algorithm, constituted by two adaptation loops, namely an inner constant throughput transmission regime, and an outer switching control regime, which jointly maintain the target bit rate of the system, while employing a set of distinct operating modes.

The financial support of the following organisations is gratefully acknowledged: EPSRC, UK in the framework of the contract GR/K 74043; the European Commission;

ICT'2000, 22-25 May, Acapulco, Mexico

Area: Joint source and channel coding

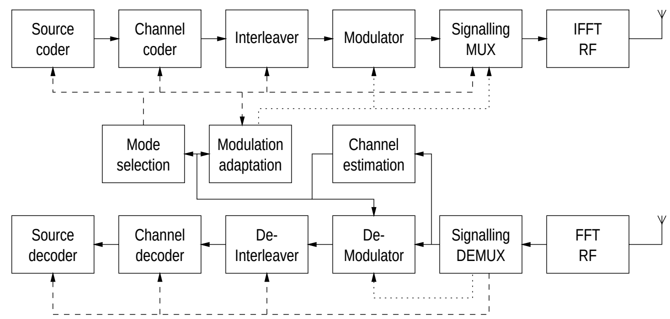

Figure 1: Schematic model of the multi–mode adaptive OFDM system

2. SYSTEM OVERVIEW

The structure of the proposed adaptive Orthogonal Frequency Division Multiplex (OFDM) transceiver is depicted in Figure 1. The transmitter chain is portrayed at the top of the diagram, which consists of the source- and channel encoders, the channel interleaver, the adaptive modulator, a multiplexer adding signalling information to the transmitted data, and an Inverse Fast Fourier Transform (IFFT) based OFDM transmitter.

The receiver seen at the bottom of the graph consists of an FFT-based demodulator and Radio-frequency (RF) receiver, a demultiplexer (DEMUX) extracting the signalling information, an adaptive demodulator, a de–interleaver/channel decoder, and the source decoder. The parameter adaptation linking the receiver- and transmitter chain consists of a channel quality estimator and the mode selection- as well as the modulation adaptation blocks.

The open–loop control structure of the adaptation algorithms can be observed in the figure, where the receiver's operation is controlled by the modem mode signalling information that is contained in or part of the received OFDM symbol, while the channel quality information estimated by the receiver is employed, in order to determine the modem mode parameter set to be employed by the transmitter. The two distinct adaptation loops distinguished by dotted and dashed lines are the inner and outer adaptation regimes, respectively. The outer adaptation loop controls the overall throughput of the system, so that a fixed–delay decoding of the received data packets becomes possible. This controls the packet size of the channel codec, the block length of the channel encoder

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

and interleaver, as well as the target throughput of the inner adaptation loop. The operation of the adaptive modulator, controlled by the inner loop, is transparent to the rest of the system. The operation of the adaptation loops is described in more detail below.

2.1. System Parameters

The transmission parameters have been adopted from the TDD–mode of the Pan-European UMTS system [5], having a carrier frequency of 1.9GHz and a TDD-frame and time slot duration of 4.615ms and 122μs, respectively. The sampling rate is assumed to be 3.78MHz, leading to a 1024–subcarrier OFDM symbol, having a cyclic extension of 64 samples in each time slot. In order to assist in the spectral shaping of the OFDM signal, there are a total of 206 virtual subcarriers at the bandwidth boundaries.

The 7-kHz bandwidth PictureTel audio codec has been chosen for this system because of its good audio quality, robustness to packet dropping and adjustable bit rate, which will be discussed in more depth below.

The channel encoder / interleaver combination is a convolutional constituent coding based turbo encoder [6] employing block interleavers with a subsequent pseudo–random channel interleaver. The constituent recursive systematic convolutional (RSC) encoders are of constraint length 3, with octal generator polynomials of $(7,5)$ and eight iterations are performed at the decoder, utilising the so-called Maximum Aposteriory (MAP) algorithm and the log-likelihood ratio soft inputs provided by the demodulator.

The channel model consists of a four path COST 207 Typical Urban impulse response [7], where each impulse is subjected to independent Rayleigh fading having a normalised Doppler frequency of $2.25 \cdot 10^{-6}$, corresponding to a pedestrian scenario with a walking speed of 3mph.

3. CONSTANT THROUGHPUT ADAPTIVE MODULATION

The constant throughput adaptive algorithm attempts to allocate the required number of bits for transmission to the specific OFDM subcarriers exhibiting a low BER, while the use of high BER subcarriers is minimised. Sub–band adaptive modulation [8], where the modulation scheme is adapted not on a subcarrier–by–subcarrier basis, but for blocks of adjacent subcarriers, is employed in order to simplify the signalling requirements.

If the impulse response of the channel $h(t, \tau)$ varies slowly compared to the OFDM symbol duration, then the Fourier transform of the impulse response during the OFDM symbol exists, and the data symbols transmitted in the subcarriers $n \in [0, \ldots, N]$ are exposed to the frequency–domain fading determined by the instantaneous channel transfer function of $H(t, n \cdot \Delta f) = H_n$. The allocation of bits to subcarriers is based on the estimated frequency domain channel transfer function $\hat{H}_n$. On the basis of this and the overall signal–to–noise ratio (SNR) γ, the local SNR in each subcarrier n can be calculated:

$$\gamma_n = \gamma / \left|\hat{H}_n\right|^2 .$$

The predicted BER $p_e(\gamma_n, m)$ in each subcarrier n and using each of the possible modulation schemes $m \in [0, \ldots, M]$ can now be computed and summed over the N_j sub–carriers in sub–band j, in order to yield the expected number of bit errors for each sub–band and for each modulation scheme:

$$e(j, m) = \sum_i p_e(\gamma_i, m)$$

for all subcarrier indices i in sub–band j. In our case, four modulation schemes are employed for $m = 0, \ldots, 3$, which are “no transmission”, BPSK, QPSK and 16-QAM, respectively. Clearly, $e(j, 0) = 0$, and the other bit error probabilities can be evaluated using the Gaussian Q-function [9]. The number of bits transmitted in sub–band j, when using modulation scheme m is denoted by $b(j, m)$.

The bit allocation regime operates iteratively, allocating bits to subcarriers by choosing the specific subcarriers for transmitting the next bit to be assigned for transmission, which increases the system's BER by the smallest amount. In other words, the bits to be transmitted are allocated consecutively, commencing by assigning bits to the highest channel quality subcarriers, gradually involving the lower channel quality carriers.

More explicitly, for each sub–band a state variable s_j is initialised to 0, and then the subband index j, for which the differential BER increment $(e_{s_j+1} - e_{s_j})/(b_{s_j+1} - b_{s_j})$ due to assigning the next bit to be transmitted is the lowest is found. The state variable s_j is incremented from 0, if it is not yet set to the index of the highest order modulation mode, ie 16QAM. This search for the lowest BER 'cost' or BER penalty, when allocating additional bits is repeated, until the total number of bits allocated to the current OFDM symbol is equal to or higher than the target number of bits to be transmitted. Clearly, the higher the target number of bits to be transmitted by each OFDM symbol, the higher the BER, since gradually lower and lower channel quality subcarriers have to be involved.

The transmitter modulates the data bits using the specific modulation schemes indexed by the state variables s_j, eventually padding the data with dummy bits, in order to maintain the required constant data throughput. The specific modulation schemes chosen for the different sub–bands have to be explicitly signalled to the receiver for accurate demodulation. Alternatively, blind subband modem mode detection algorithms can be employed at the receiver [2]. For the scope of this paper, we assume 32 sub–bands of 32 subcarriers in each 1024-subcarrier OFDM symbol. Perfect channel estimation and subband modem mode signalling were assumed.

3.1. Performance

Figure 2 shows an example of the fixed throughput adaptive OFDM modulation scheme's performance under the channel conditions characterised above, for a block length of 578 coded bits per OFDM symbol. As a comparison, a fixed BPSK modem transmitting the same number of bits in the same channel, employing 578 out of 1024 subcarriers, is depicted. The number of bits per OFDM symbol is based on a 200-bit useful data throughput, which corresponds to a 10 kbps useful data rate, padded with 89 bits, which can contain a check–sum for error detection and high–level signalling information. Furthermore, half–rate channel coding was used.

The BER plotted in the figure is the hard decision based bit error rate at the receiver before channel decoding. It can be seen that the adaptive modulation yields a significantly improved performance, which is reflected also in the associated Frame Error Rate (FER). This FER was defined as the probability of a decoded block containing errors, in which case it is unusable for the audio source decoder and hence it is dropped.

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

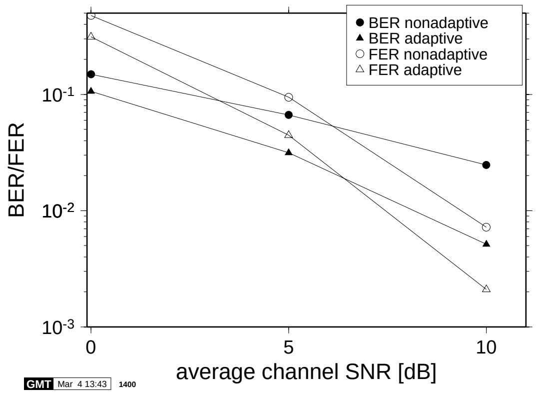

Figure 2: BER and FER for the fixed throughput adaptive (NoTX, BPSK, 4QAM and 16QAM) scheme and for the fixed-mode BPSK scheme in fading time dispersive channel for a block length of 578 coded bits.

This error event can be detected by using the check–sum in the data symbol.

4. MULTI–MODE ADAPTATION

While the fixed throughput adaptive algorithm described above copes well with the *frequency–domain* fading of the channel, there is also a medium-term *time-domain variation* of the overall channel capacity. Hence - in addition to the previously proposed fixed-rate frequency-domain bitallocation scheme - in this section we propose the employment of a time-variant bitrate scheme, in order to gauge its additional performance potential benchmarked against the fixed-rate schemes. We will then also contrive appropriate matching audio transceivers at a later stage. However, our experience demonstrated that it was an arduous task to employ powerful block–based turbo channel coding schemes in conjunction with variable throughput adaptive schemes for real–time applications, such as voice or video telephony. Nonetheless, a multi–mode adaptive system can be designed that allows us to switch between a set of different source– and channel coders as well as transmission parameters, depending on the overall instantaneous channel quality. We have investigated the employment of the estimated overall BER at the output of the receiver averaged over an OFDM symbol, which is the sum of all the $e(j, s_j)$ subband BER contributions after adaptation. On the basis of this expected input BER of the channel decoder, the probability of a frame error (FE) must be estimated, and compared with the expected FER of the other modem modes. Then, the mode having the lowest FER is selected and the source coder, the channel coder and the adaptive modem are set up accordingly. At this stage we note that it is not always the most robust modem mode that results in the lowest FEC-coded FER, since upon transmitting more bits per OFDM symbol the turbo channel codec becomes more powerful due to its longer channel interleaver.

We have defined four different operating modes, which correspond to the uncoded audio data rates of 10, 16, 24, and 32 kbps at the source encoder's output. With half–rate channel coding and allowing for check–sum and signalling overheads, the number of transmitted coded bits per OFDM symbol is 578, 722, 1058, and 1458 for the four source-coded modes, respectively.

4.1. Mode switching

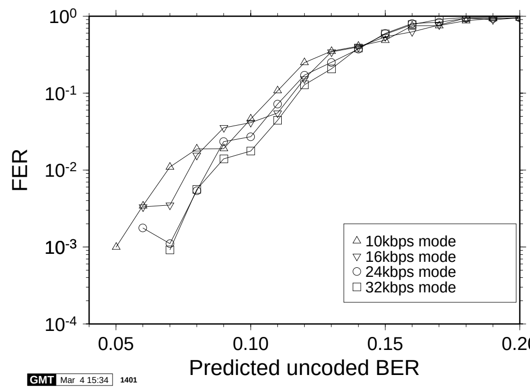

Figure 3: Frame error rate versus the predicted uncoded BER for 10kbps, 16kbps, 24kbps and 32kbps modes.

Figure 3 shows the observed FER for all four modes versus the uncoded BER that was predicted at the transmitter. The predicted BER was discretised into intervals of 1%, and the FER was averaged over these intervals. It can be seen that for estimated BER values below 5% no frame errors were observed - irrespective of the source-coded operational modes. Observe in the figure that at higher estimated BER values the higher bitrate modes - such as 24 and 32 kbps - exhibited a lower FER than the lower bitrate modes, which was due to the turbo codec's improved performance in conjunction with longer interleaving block lengths. For example, a FER of 1% was observed for a 7% predicted input error rate for the 10kbps mode, while 8% to 9% predicted uncoded BERs were tolerable for maintaining the same FER in conjunction with the longer interleved blocks associated with the 24 and 32 kbps audio modes.

In this contribution we assumed the best–case scenario of invoking the experimentally determined FER statistics of Figure 3 for the mode switching algorithm. In this case, the FERs corresponding to the predicted overall BER values for the different modes were compared, and the mode exhibiting the lowest FER was chosen for transmission.

5. THE PICTURETEL CODEC

The PictureTel Transform Coder (PTC) is one of the candidates for the new ITU-T wideband audio coding standard. It is based on the so-called Modulated Lapped Transform (MLT)[11], followed by a quantisation stage using a perceptually motivated psychoacoustic quantisation model and Huffman coding for encoding the frequency domain coefficients.

At its input the PTC expects frames of 320 Pulse Code Modulated (PCM) audio samples, obtained by sampling an audio signal at a frequency of 16 kHz with a quantiser resolution of 14, 15 or 16 bit. Furthermore, the input samples are assumed only to contain frequency components up to 7 kHz. The PictureTel standard proposal recommends operating the codec at output bitrates of 16 kbps, 24 kbps or 32 kbps, generating output frame lengths of 320, 480 or 640 bits per 20ms, respectively, for which the codec was optimized. The delay encountered by an audio frame, when passing through the codec (consisting of encoder and decoder) can be estimated to be on the order of about 60 ms, which is a result of the time

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

domain frame overlapping technique and the computational delay inherent in the codec.

Since the PTC employs Huffman coding for encoding the frequency domain MLT coefficients, the decoding is very sensitive to bit errors. Hence a single bit error can render the whole audio frame undecodable. The PTC's standard reaction to such a frame error is simply to repeat the previous frame of coefficients, with the result that such error events are nearly imperceivable by the user, as long as they occur relatively rarely. For bursts of frame errors the output signal is gradually muted after decoding the first erroneous frame, which is distinctively audible due to the relatively large frame size of 20 ms.

6. SIMULATION RESULTS

Our discussions related to the associated design tradeoffs and the impact of an automatic bit rate selection scheme on the audio quality of the system will be mainly based on measurements performed around channel SNR values of 5 dB, since for very low SNRs of around 0 dB the frame dropping rate or FER is excessive, yielding an unacceptable audio quality. By contrast, for high channel SNRs around 10 dB the FER is too low for enabling us to illustrate the tradeoffs between the audio quality and FER effectively. A tentative estimate of the average quality of the reconstructed audio signal is provided by audio segmental SNR calculations, which provide an approximate measure of the subjectively perceived audio quality [12], especially in the presence of frame dropping in conjunction with perceptual masking based transform assisted audio coding.

6.1. Frame Error Results

The basic tradeoff between the system's throughput and audio frame dropping-rate is illustrated for the 5 dB channel SNR scenario with the aid of our four fixed bit rate modes in Table 1. The first column reflects for each bit rate the associated frame dropping rate or FER that we will encounter.

As expected, by increasing the throughput, the frame dropping rate will also increase, since a high proportion of reduced-quality subcarriers has to be used for conveying the increased number of audio bits, although the performance of the turbo channel codec improves. Experiments have shown that a frame dropping rate or FER of around 5% in conjunction with the 16 kbps fixed bit rate mode is still sufficiently low, in order to provide a perceptually acceptable audio quality. In the second to fourth column of Table 1, the relative frequency of encountering error-free audio frames for the different audio bit rates is portrayed. Observe furthermore in the table that the same performance figures were also summarised for two different transmission schemes denoted by *Switch I* and *Switch II*. These schemes invoked a system philosophy allowing the bitrate to become time-variant and controlling the audio source codec and channel codec on a time-variant basis, in order to take this time-variant behaviour into account, as it will be highlighted below.

Specifically, both of our experimental switching regimes, namely *Switch I* and *Switch II* employed the switching algorithm described in Section 4.1, with the only difference that Switch I incorporated in addition to the three standard bit rates of 16, 24 and 32 kbps - which were proposed by the PictureTel company - a 10 kbps mode, with the intention of lowering the frame dropping rate further due to the more modest 'loading' of the OFDM symbols. For these switching schemes the 5 dB SNR related results in Table 1 underline that for example in comparison to the 16 kbps mode the system throughput was very much improved, conveying most of the audio frames in the 24 and 32 kbps mode, rather than in the 16 kbps mode, while maintaining the same frame dropping rate of 5.58 %. Although exhibiting a slightly lower frame dropping rate, the Switch I scheme was shown to produce an audio quality inferior to that of the Switch II scheme. This was due to the employment of the 10 kbps bit rate mode in the Switch I scheme, which produced a relatively low subjective audio quality. In this context it is interesting to see that although the Switch I scheme assigns about 22% percent of all frames to the 10 kbps mode, the frame dropping rate was increased only by about 1.1 %, when disabling this mode in the Switch II scheme. This is an indication of the conservative decision regime of our bit rate selector. The relative frequency of invoking the different bit rates in conjunction with the Switch II scheme has been evaluated additionally for channel SNRs of 0 dB and 10 dB, which characterises the operation of the bit rate selector once again. The associated results are presented in Table 2.

6.2. Audio Segmental SNR Performance

In addition to our previous results Figure 4 displays the cumulative density function (CDF) of the segmental SNR (SEGSNR) obtained from the reconstructed signal of an audio test sound at the output of the PTC decoder for the schemes described above at a channel SNR of 5 dB. The step-function like CDF increase at a SEGSNR of 0 dB is equal to the frame dropping rate of the corresponding scheme, given in Table 2. As expected, the fixed mode 32 kbps scheme performs best, when neglecting frame drops, since the slope of the CDF or the corresponding PDF exhibits the lowest values of all schemes at low SEGSNRs and the highest values at high SEGSNRs. The fixed 32 kbps scheme exhibits a similar performance to the fixed 24 kbps scheme, while the fixed 10 kbps scheme exhibits the lowest performance. This is, why this scheme was excluded in the Switch II arrangement. However, FERs in excess of 10% can result in distinctively audible artifacts, which virtually renders the fixed 24 kbps and 32 kbps modes unacceptable. *Hence, our proposed switching scheme, which is based on the* 16, 24 *and* 32 *kbps standard bit rates, achieves at medium SNRs the best compromise between average audio quality and frame dropping rate, which has been verified by our listening tests.*

7. FURTHER WORK

As outlined in Section 4.1, the mode switching algorithm operates on the basis of statistically evaluated experimental results for the prediction of the FER. A robust, channel-independent switching regime on the basis of the turbo coder's channel quality perceptions can overcome this dependence. Furthermore, a target–FER driven switching scheme instead of the minimal–FER algorithm employed for this series of experiments will be investigated in future.

8. REFERENCES

[1] W.T. Webb and R. Steele, "Variable Rate QAM for mobile radio," *IEEE Transactions on Communications*, vol. 43, pp. 2223 – 2230, July 1995.

[2] L. Hanzo, W.T. Webb, T. Keller: Single- and Multi-carrier Quadrature Amplitude Modulation: Principles and Applications for Personal Communications, WATM and Broadcasting; IEEE Press-John Wiley, 2000

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000

Scheme	FER[%]	Rel.Fr.: 10kbps [%]	Rel.Fr.: 16kbps [%]	Rel.Fr.: 24kbps [%]	Rel.Fr.: 32kbps [%]
Fixed-10kbps	4.45	95.55	0.0	0.0	0.0
Fixed-16kbps	5.58	0.0	94.42	0.0	0.0
Fixed-24kbps	10.28	0.0	0.0	89.72	0.0
Fixed-32kbps	18.65	0.0	0.0	0.0	81.35
Switch-I	4.44	21.87	13.90	11.59	48.20
Switch-II	5.58	0.0	34.63	11.59	48.20

Table 1: FER and relative frequency of different bitrates in the fixed bit rate- and the switching schemes for a channel SNR of 5 dB

Ch. SNR [dB]	FER [%]	Rel.Fr.: 10kbps [%]	Rel.Fr.: 16kbps [%]	Rel.Fr.: 24kbps [%]	Rel.Fr.: 32kbps [%]
0	37.69	0.0	37.79	14.42	10.10
5	5.58	0.0	34.63	11.59	48.20
10	0.34	0.0	7.81	5.61	86.24

Table 2: FER and relative frequency of different bit rates in the Switch II scheme for a channel SNR's of 0, 5 and 10 dB

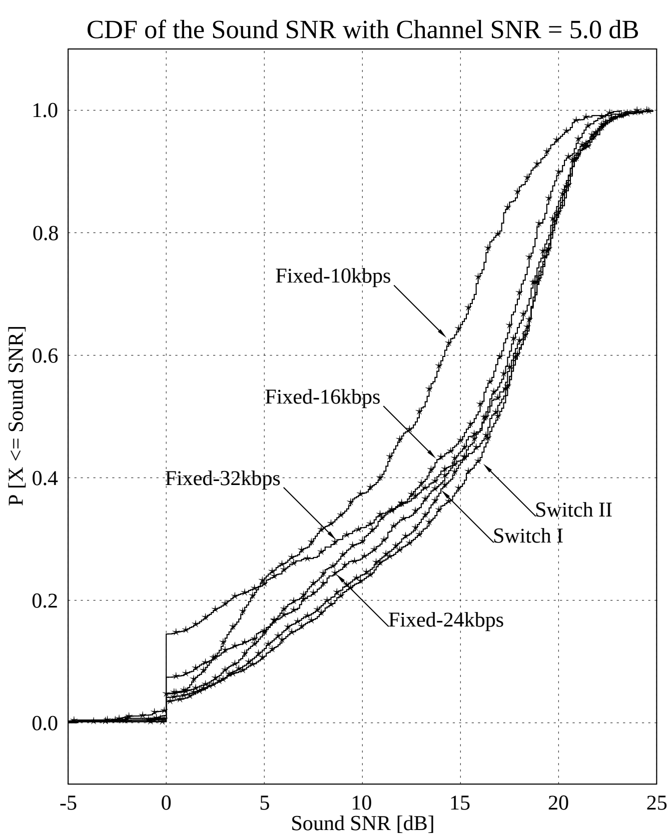

Figure 4: Typical CDF of the Segmental SNR of a reconstructed audio signal transmitted over the COST 207 fading time dispersive channel at 3mph.

[3] H. Matsuako, S. Sampei, N. Morinaga and Y. Kamio, "Adaptive Modulation systems with variable coding rate concatenated code for high quality multi-media communication systems," in *Proceedings of IEEE Vehicular Technology Conference*, pp. 487 – 491, 1996.

[4] A. J. Goldsmith and S. G. Chua, "Variable Rate Variable Power MQAM for Fading Channels," *IEEE Transactions on Communications*, vol. 45, pp. 1218 – 1230, October 1997.

[5] E. Dahlman, B. Gudmundson, M. Nilsson and J. Skold, " UMTS and IMT-2000 based on wideband CDMA," *IEEE Commun. Mag.*, Vol. 36, No. 9, pp. 70-80, Sept. 1998.

[6] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near shannon limit error-correcting coding and decoding: Turbo-codes," in *Proc. ICC*, pp. 1064–1070, 1993.

[7] M. Failli, "Digital land mobile radio communications cost 207," tech. rep., European Commission, 1989.

[8] T. Keller and L. Hanzo, "Adaptive orthogonal frequency division multiplexing schemes," in *Proc. ACST Summit*, (Rhodos, Greece), pp. 794–799, June 1998.

[9] J. G. Proakis, *Digital Communications*. London: McGraw–Hill, 2^{nd} ed., 1989.

[10] T. Keller, L. Hanzo: Blind-detection Assisted Sub-band Adaptive Turbo-Coded OFDM Schemes, Proc. of VTC'99, Houston, USA, May, 1999, pp 489-493

[11] H. S. Malvar, *Signal Processing with Lapped Transforms*. Boston, MA: Artech House, 1992.

[12] N. Kitawaki, M. Honda, and K. Itoh, "Speech-Quality Assessment Methods for Speech-Coding Systems," *IEEE Comms. Magazine*, vol. 22, pp. 26–33, October 1984.

ISBN 968-36-7763-0 / 22-25 May 2000 © ICT 2000