Turbo-Detected Unequal Protection MPEG-4 Audio Transceiver Using Convolutional Codes, Trellis Coded Modulation and Space-Time Trellis Coding

N. S. Othman, S. X. Ng and L. Hanzo

School of ECS, University of Southampton, SO17 1BJ, UK.
Tel: +44-23-8059 3125, Fax: +44-23-8059 4508
Email: {nso01r,sxn,lh}@ecs.soton.ac.uk, http://www-mobile.ecs.soton.ac.uk

Abstract – **A jointly optimised turbo transceiver capable of providing unequal error protection is proposed for employment in an MPEG-4 aided audio transceiver. The transceiver advocated consists of Space-Time Trellis Coding (STTC), Trellis Coded Modulation (TCM) and two different-rate Non-Systematic Convolutional codes (NSCs) used for unequal error protection. A benchmarker scheme combining STTC and a single-class protection NSC is used for comparison with the proposed scheme. The audio performance of the both schemes is evaluated when communicating over uncorrelated Rayleigh fading channels. It was found that the proposed unequal protection turbo-transceiver scheme requires about two dBs lower transmit power than the single-class turbo benchmarker scheme in the context of the MPEG-4 audio transceiver, when aiming for an effective throughput of 2 bits/symbol, while exhibiting a similar decoding complexity.**

1. MOTIVATION AND BACKGROUND

The MPEG-4 standard [1, 2] defines a comprehensive multimedia content representation scheme that is capable of supporting numerous applications - such as streaming multimedia signals over the internet/intranet, content-based storage and retrieval, digital multimedia broadcast or mobile communications. The audio-related section of the MPEG-4 standard [3] defines audio codecs covering a wide variety applications - ranging from narrowband low-rate speech to high quality multichannel audio, and from natural sound to synthesized sound effects as a benefit of its object-based approach used for representing the audio signals.

The MPEG-4 General Audio (GA) encoder is capable of compressing arbitrary natural audio signals. One of the key components of the MPEG-4 GA encoder is the Time/Frequency (T/F) compression scheme constituted by the Advanced Audio Coding (AAC) and Transform based Weighted Vector Quantization (TwinVQ), which is capable of operating at bitrates ranging from 6 kbit/s to broadcast quality audio at 64 kbit/s [1].

The financial support of the EPSRC, Swindon UK and the EU under the auspices of the PHOENIX, NEWCOM and NEXWAY projects is gratefully acknowledged.

The MPEG-4 T/F codec is based on the MPEG-2 AAC standard, extended by a number of additional functionalities, such as Perceptual Noise Substitution (PNS) and Long Term Prediction (LTP) for enhancing the achievable compression performance, and combined with the TwinVQ for operation at extremely low bit rates. Another important feature of this codec is its robustness against transmission errors in error-prone propagation channels [4]. The error resilience of the MPEG-4 T/F codec is mainly attributed to the so-called Virtual Codebook tool (VCB11), Reversible Variable Length Coding tool (RVLC) and Huffman Codeword Reordering tool (HCR) [4, 5], which facilitate the integration of the MPEG-4 T/F codec into wireless systems.

In this study the MPEG-4 audio codec was incorporated in a sophisticated unequal-protection turbo transceiver using joint coding and modulation as inner coding, twin-class convolutional outer coding as well as space time coding based spatial diversity. Specifically, maximal minimum distance Non-Systematic Convolutional codes (NSCs) [10, p. 331] having two different code-rates were used as outer encoders for providing unequal audio protection. On one hand, Trellis Coded Modulation (TCM) [6–8] constitutes a bandwidth-efficient joint channel coding and modulation scheme, which was originally designed for transmission over Additive White Gaussian Noise (AWGN) channels. On the other hand, Space-Time Trellis Coding (STTC) [7, 9] employing multiple transmit and receive antennas is capable of providing spatial diversity gain. When the spatial diversity order is sufficiently high, the channel's Rayleigh fading envelope is transformed to a Gaussian-like near-constant envelope. Hence, the benefits of a TCM scheme designed for AWGN channels will be efficiently exploited, when TCM is concatenated with STTC.

We will demonstrate that significant iteration gains are attained with the aid of the proposed turbo transceiver. The paper is structured as follows. In Section 2 we describe the MPEG-4 audio codec, while in Section 3 the architecture of the turbo transceiver is described. We elaborate further by characterising the achievable system performance in Section 4 and conclude in Section 5.

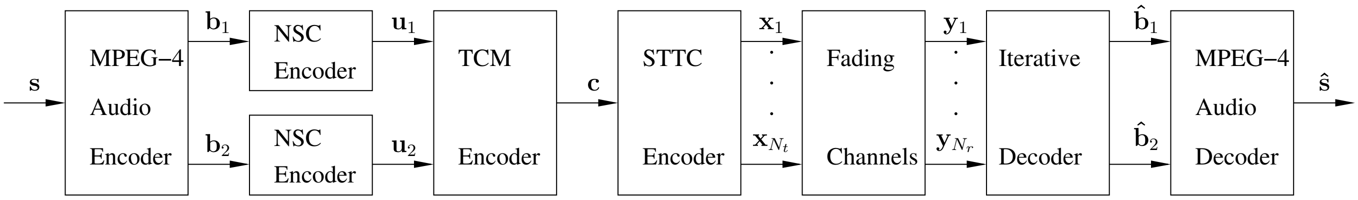

Figure 1: Block diagram of the serially concatenated STTC-TCM-2NSC assisted MPEG-4 audio scheme. The notations $\mathbf{s}$, $\hat{\mathbf{s}}$, $\mathbf{b}_i$, $\hat{\mathbf{b}}_\mathbf{i}$, $\mathbf{u}_i$, $\mathbf{c}$, $\mathbf{x}_j$ and $\mathbf{y}_k$ denote the vector of the audio source symbol, the estimate of the audio source symbol, the class-i audio bits, the estimates of the class-i audio bits, the encoded bits of class-i NSC encoders, the TCM coded symbols, the STTC coded symbols for transmitter j and the received symbols at receiver k, respectively. Furthermore, N_t and N_r denote the number of transmitters and receivers, respectively. The symbol-based channel interleaver between the STTC and TCM schemes as well as the two bit-based interleavers at the output of NSC encoders are not shown for simplicity. The iterative decoder seen at the right is detailed in Figure 2.

2. AUDIO SYSTEM OVERVIEW

As mentioned above, the MPEG-4 AAC is based on time/frequency audio coding, which provides redundancy reduction by exploiting the correlation between subsequent audio samples of the input signal. Furthermore, the codec uses perceptual modelling of the human auditory system for masking the quantisation distortion of the encoded audio signals by allowing more distortion in those frequency bands, where the signal exhibits higher energy peaks and vice versa [4, 5].

The MPEG-4 AAC is capable of providing an attractive audio quality versus bitrate performance, yielding high-fidelity audio reconstruction for bit rates in excess of 32 kbit/s per channel. In the proposed wireless system the MPEG-4 AAC is used for encoding the stereo audio file at a bit rate of 48 kbit/s. The audio input signal was sampled at 44.1 kHz and hence results in an audio framelength of 23.22 ms, which corresponds to 1024 audio input samples. The compressed audio information is formatted into a packetized bitstream, which conveyed one audio frame. In our system, the average transmission frame size is approximately 1116 bits per frame. The audio Segmental Signal to Noise Ratio (SegSNR) of this configuration was found to be S_0 = 16.28dB, which gives a transparent audio quality.

It is well recognised that in highly compressed audio bitstreams a low bit error ratio (BER) may lead to perceptually unacceptable distortion. In order to prevent the complete loss of transmitted audio frames owing to catastrophic error propagation, the most sensitive bits have to be well protected from channel errors. Hence, in the advocated system Unequal Error Protection (UEP) is employed, where the compressed audio bitstream was partitioned into two sensitivity classes. More explicitly, an audio bit, which resulted in a SegSNR degradation above 16 dB upon its corruption was classified into protection class-1. A range of different audio files were used in our work and the results provided are related to a 60 seconds long excerpt of Mozart's "Clarinet Concerto (2nd movement - Adagio)". From the bit sensitivity studies using this audio file as the source, we found that approximately 50% of the total number of MPEG-4 encoded bits falls into class-1.

At the receiver, the output of the turbo transceiver is decoded using the MPEG-4 AAC decoder. During the decoding process, the erroneously received audio frames were dropped and replaced by the previous error-free audio frame for the sake of avoiding an even more dramatic error-infested audio-quality degradation [11, 12].

3. THE TURBO TRANSCEIVER

The block diagram of the serially concatenated STTC-TCM-2NSC turbo scheme using a STTC, a TCM and two different-rate NSCs as its constituent codes is depicted in Figure 1. Since the number of class-1 audio bits is approximately the same as that of the class-2 audio bits and there are approximately 1116 bits per audio frame, we protect the 558-bit class-1 audio sequence using a rate-R_1 NSC encoder and the 558-bit class-2 sequence using a rate-R_2 NSC encoder. Let us denote the turbo scheme as STTC-TCM-2NSC-1 when the NSC coding rates of $R_1 = k_1/n_1 = 1/2$ and $R_2 = k_2/n_2 = 3/4$ are used. Furthermore, when the NSC coding rates of $R_1 = 2/3$ and $R_2 = 3/4$ are used, we denote the turbo scheme as STTC-TCM-2NSC-2. The code memory of the class-1 and class-2 NSC encoders is $L_1 = 3$ and $L_2 = 3$, respectively. The class-1 and class-2 NSC coded bit sequences are interleaved by two separate bit interleavers, before they are fed to the rate-$R_3 = 3/4$ TCM [6–8] scheme having a code memory of $L_3 = 3$. Code termination was employed for the NSCs, TCM [6–8] and STTC codecs [7, 9]. The TCM symbol sequence is then symbol-interleaved and fed to the STTC encoder. We invoke a 16-state STTC scheme having a code memory of $L_4 = 4$ and $N_t = 2$ transmit antennas, employing $M = 16$-level Quadrature Amplitude Modulation (16QAM) [8]. The STTC employing $N_t = 2$ requires one 16QAM-based termination symbol. The overall coding rate is given by $R_{s1} = 1116/2520 \approx 0.4429$ and $R_{s2} = 1116/2152 \approx 0.5186$ for the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes, respectively. The effective throughput of the STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes is $\log_2(M)R_{s1} \approx 1.77$ Bits Per Symbol

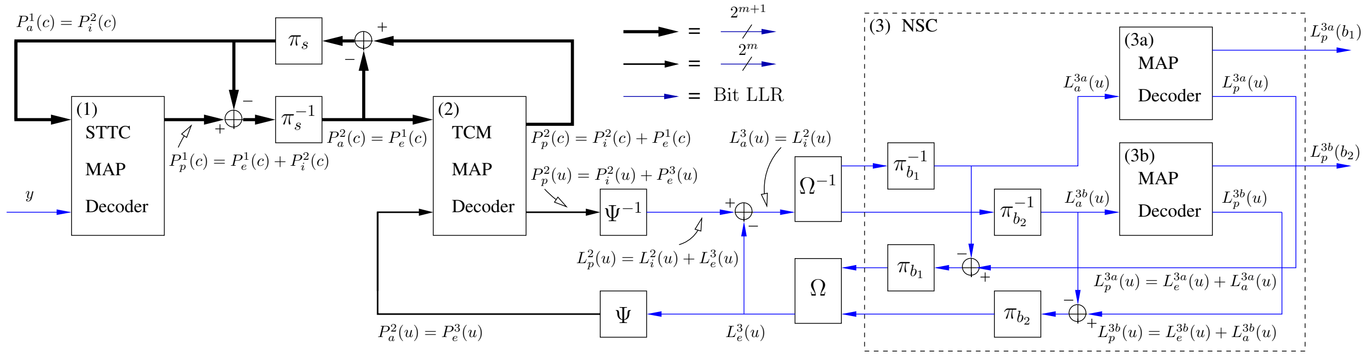

Figure 2: Block diagram of the STTC-TCM-2NSC turbo detection scheme seen at the right of Figure 1. The notations $\pi_{(s,b_i)}$ and $\pi^{-1}_{(s,b_i)}$ denote the interleaver and deinterleaver, while the subscript s denotes the symbol-based interleaver of TCM and the subscript b_i denotes the bit-based interleaver for class-i NSC. Furthermore, Ψ and Ψ^{-1} denote LLR-to-symbol probability and symbol probability-to-LLR conversion, while Ω and Ω^{-1} denote the parallel-to-serial and serial-to-parallel converter, respectively. The notation m denotes the number of information bits per TCM coded symbol. The thickness of the connecting lines indicates the number of non-binary symbol probabilities spanning from a single LLR per bit to 2^m and 2^{m+1} probabilities [13] ©IEE, 2004, Ng, Chung and Hanzo.

(BPS) and $\log_2(M)R_{s2} \approx 2.07$ BPS, respectively.

At the receiver, we employ $N_r = 2$ receive antennas and the received signals are fed to the iterative decoders for the sake of estimating the audio bit sequences in both class-1 and class-2, as seen in Figure 1. The STTC-TCM-2NSC scheme's turbo decoder structure is illustrated in Figure 2, where there are four constituent decoders, each labelled with a round-bracketed index. The Maximum A-Posteriori (MAP) algorithm [7] operating in the logarithmic-domain are employed by the STTC, TCM and the two NSC decoders, respectively. The notations $P(.)$ and $L(.)$ in Figure 2 denote the logarithmic-domain symbol probabilities and the Logarithmic-Likelihood Ratio (LLR) of the bit probabilities, respectively. The notations c, u and b_i in the round brackets $(.)$ in Figure 2 denote TCM coded symbols, TCM information symbols and the class-i audio bits, respectively. The specific nature of the probabilities and LLRs is represented by the subscripts a, p, e and i, which denote $a\ priori$, $a\ posteriori$, $extrinsic$ and $intrinsic$ information, respectively. The probabilities and LLRs associated with one of the four constituent decoders having a label of $\{1, 2, 3a, 3b\}$ are differentiated by the identical superscripts of $\{1, 2, 3a, 3b\}$. Note that the superscript 3 is used for representing the two NSC decoders of $3a$ and $3b$. The iterative turbo-detection scheme shown in Figure 2 enables an efficient information exchange between STTC, TCM and NSCs constituent codes for the sake of achieving spatial diversity gain, coding gain, unequal error proctection and a near-channel-capacity performance. The information exchange mechanism between each constituent decoders is detailed in [13].

For the sake of benchmarking the scheme advocated, we created a powerful benchmark scheme by replacing the TCM and NSC encoders of Figure 1 by a single NSC codec having a coding rate of $R_0 = k_0/n_0 = 1/2$ and a code memory of $L_0 = 6$. We will refer to this benchmarker scheme as the STTC-NSC arrangement. All audio bits are equally protected in the benchmarker scheme by a single NSC encoder and a STTC encoder. A bit-based channel interleaver is inserted between the NSC encoder and STTC encoder. Taking into account the bits required for code termination, the number of output bits of the NSC encoder is $(1116 + k_0L_0)/R_0 = 2244$, which corresponds to 561 16QAM symbols. Again, a 16-state STTC scheme having $N_t = 2$ transmit antennas is employed. After code termination, we have $561 + 1 = 562$ 16QAM symbols or $4(562) = 2248$ bits in a transmission frame at each transmit antenna. The overall coding rate is given by $R = 1116/2248 \approx 0.4964$ and the effective throughput is $\log_2(16)R \approx 1.99$ BPS, both of which are very close to the corresponding values of the STTC-TCM-2NSC-2 scheme. A decoding iteration of the STTC-NSC benchmarker scheme is comprised of a STTC decoding and a NSC decoding step.

We will quantify the decoding complexity of the proposed STTC-TCM-2NSC scheme and that of the benchmarker scheme using the number of decoding trellis states. The total number of decoding trellis states per iteration for the proposed scheme employing 2 NSC decoders having a code memory of $L_1 = L_2 = 3$, TCM having $L_3 = 3$ and STTC having $L_4 = 4$, is given by $S = 2^{L_1} + 2^{L_2} + 2^{L_3} + 2^{L_4} = 40$. By contrast, the total number of decoding trellis states per iteration for the benchmarker scheme having a code memory of $L_0 = 6$ and STTC having $L_4 = 4$, is given by $S = 2^{L_0} + 2^{L_4} = 80$. Therefore, the complexity of the proposed STTC-TCM-2NSC scheme having two iterations is equivalent to that of the benchmarker scheme having a single iteration, which corresponds to 80 decoding states.

4. SIMULATION RESULTS

In this section we evaluate the performance of the proposed MPEG-4 based audio telephone schemes using both the Bit Error Ratio (BER) and the Segmental Signal to Noise Ratio (SegSNR).

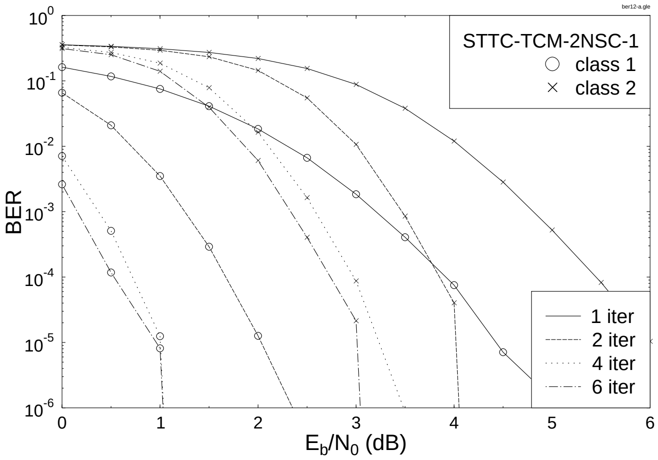

Figure 3: BER versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC-1 assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.77 BPS**.

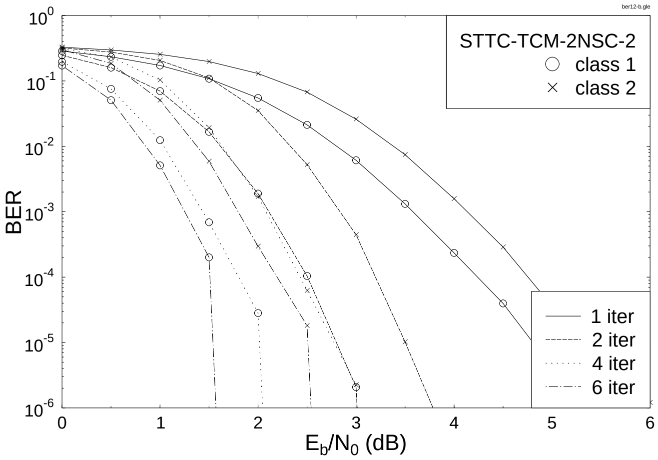

Figure 4: BER versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC-2 assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **2.07 BPS**.

Figures 3 and 4 depict the BER versus Signal to Noise Ratio (SNR) per bit, namely E_b/N_0, performance of the 16QAM-based STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 schemes, respectively, when communicating over uncorrelated Rayleigh fading channels. As we can observe from Figures 3 and 4, the gap between the BER performance of the class-1 and class-2 audio bits is wider for STTC-TCM-2NSC-1 compared to the STTC-TCM-2NSC-2 scheme. More explicitly, the class-1 au-

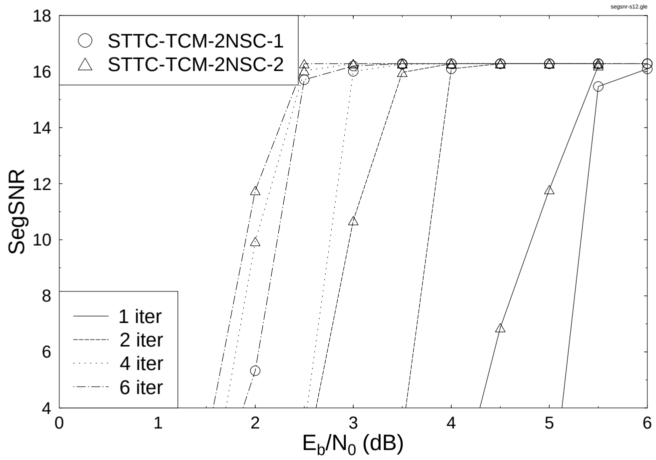

Figure 5: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-TCM-2NSC assisted MPEG-4 audio scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput of STTC-TCM-2NSC-1 and STTC-TCM-2NSC-2 was **1.77** and **2.07 BPS**, respectively.

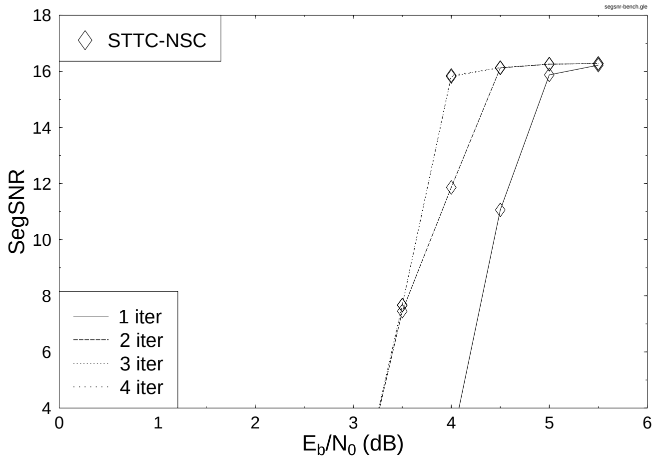

Figure 6: Average SegSNR versus E_b/N_0 performance of the 16QAM-based STTC-NSC assisted MPEG-4 audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. The effective throughput was **1.99 BPS**.

dio bits of STTC-TCM-2NSC-1 have a higher protection at the cost of a lower throughput compared to the STTC-TCM-2NSC-2 scheme. However, the BER performance of the class-2 audio bits of the STTC-TCM-2NSC-1 arrangement is approximately 0.5 dB poorer than that of STTC-TCM-2NSC-2 at BER=10^{-5}.

Let us now study the audio SegSNR performance of the schemes in Figures 5 and 6. As we can see from Figure 5, the SegSNR performance of STTC-TCM-2NSC-1 is inferior in comparison to that of STTC-TCM-2NSC-2, despite providing a higher protection for the class-1 audio bits. More explicitly, STTC-TCM-2NSC-2 requires E_b/N_0 = 2.5 dB, while STTC-

TCM-2NSC-1 requires E_b/N_0 = 3 dB, when having an audio SegSNR in excess of 16 dB after the fourth turbo iteration. Hence the audio SegSNR performance of STTC-TCM-2NSC-1 is 0.5 dB poorer than that of STTC-TCM-2NSC-2 after the fourth iteration. Note that the BER of the class-1 and class-2 audio bits for the corresponding values of E_b/N_0, SegSNR and iteration index is less than 10^{-7} and 10^{-4}, respectively, for the two different turbo schemes. After the sixth iteration, the SegSNR performance of both turbo schemes becomes quite similar since the corresponding BER is low. These results demonstrate that the MPEG-4 audio decoder requires a very low BER for both class-1 and class-2 audio bits, when aiming for a SegSNR above 16 dB. In this context it is worth mentioning that Recursive Systematic Convolutional codes (RSCs) [6, 7, 10] are capable of achieving a higher iteration gain, but suffer from an error floor. Owing to this reason the SegSNR performance of the schemes employing RSCs instead of NSCs was found to be poorer. The SegSNR results of the turbo schemes employing RSCs instead of NSCs as the outer code were not shown here for reasons of space economy.

Figure 6 portrays the SegSNR versus E_b/N_0 performance of the STTC-NSC audio benchmarker scheme, when communicating over uncorrelated Rayleigh fading channels. Note that if we reduce the code memory of the NSC constituent code of the STTC-NSC benchmarker arrangement from L_0=6 to 3, the achievable performance becomes poorer, as expected. If we increased L_0 from 6 to 7 (or higher), the decoding complexity would increase significantly, while the attainable best possible performance is only marginally increased. Hence, the STTC-NSC scheme having L_0=6 constitutes a good benchmarker scheme in terms of its performance versus complexity tradeoffs. It is shown in Figures 5 and 6 that the first iteration based performance of the STTC-NSC benchmarker scheme is better than that of the proposed STTC-TCM-2NSC arrangements. However, at the same decoding complexity of 160 (240) trellis decoding states STTC-TCM-2NSC-2 having 4 (6) iterations performs approximately 2 (1.5) dB better than the STTC-NSC arrangement having 2 (3) iterations.

It is worth mentioning that other joint coding and modulation schemes directly designed for fading channels, such as for example Bit Interleaved Coded Modulation (BICM) [7, 8, 14] were outperformed by the TCM-based scheme, since the STTC arrangement rendered the error statistics more Gaussian-like [15].

5. CONCLUSIONS

In conclusion, a jointly optimised audio source-coding, outer unequal protection NSC channel-coding, inner TCM and spatial diversity aided STTC turbo transceiver was proposed for employment in a MPEG-4 wireless audio transceiver. With the aid of two different-rate NSCs the audio bits were protected differently according to their error sensitivity. The employment of TCM improved the bandwidth efficiency of the system and by utilising STTC spatial diversity was attained. The performance of the proposed STTC-TCM-2NSC scheme was enhanced with the advent of an efficient iterative joint decoding structure. The MPEG-4 audio decoder was found to require a very low BER for both classes of audio bits in order to attain a perceptually pleasing, artefact-free audio quality. The proposed twin-class STTC-TCM-2NSC scheme performs approximately 2 dB better in terms of the required E_b/N_0 than the single-class STTC-NSC audio benchmarker.

6. REFERENCES

[1] R. Koenen , “MPEG-4 Overview,” in *ISO/IEC JTC1/SC29/WG11 N4668, version 21-Jeju Version, ISO/IEC*, (http://www.chiariglione.org/mpeg/standards/mpeg-4/mpeg-4.htm), March 2002.

[2] R. Koenen, “MPEG-4 Multimedia for Our Time,” vol. 36, pp. 26–33, February 1999.

[3] ISO/IEC JTC1/SC29/WG11 N2503, “Information Technology-Very Low Bitrate Audio-Visual Coding,” in *ISO/IEC 14496-3. Final Draft International Standard. Part 3: Audio*, 1998.

[4] J. Herre, and B. Grill, “Overview of MPEG-4 Audio and its Applications in Mobile Communications,” vol. 1, pp. 11–20, August 2000.

[5] F. Pereira and T. Ebrahimi , *The MPEG-4 Book*. New Jersey, USA: Prentice Hall PTR IMSC Press, 2002.

[6] G. Ungerböck, “Channel Coding with Multilevel/Phase Signals,” *IEEE Transactions on Information Theory*, vol. 28, pp. 55–67, January 1982.

[7] L. Hanzo, T. H. Liew and B. L. Yeap, *Turbo Coding, Turbo Equalisation and Space Time Coding for Transmission over Wireless channels*. New York, USA: John Wiley IEEE Press, 2002.

[8] L. Hanzo, S. X. Ng, W. T. Webb and T. Keller, *Quadrature Amplitude Modulation: From Basics to Adaptive Trellis-Coded, Turbo-Equalised and Space-Time Coded OFDM, CDMA and MC-CDMA Systems*. New York, USA: John Wiley IEEE Press, 2004.

[9] V. Tarokh, N. Seshadri and A. R. Calderbank, “Space-time Codes for High Rate Wireless Communication: Performance analysis and code construction,” *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[10] S. Lin and D. J. Costello, Jr, *Error Control Coding: Fundamentals and Applications*. Inc. Englewood Cliffs, New Jersey 07632: Prentice-Hall, 1983.

[11] L. Hanzo, P.J. Cherriman and J. Street, *Wireless Video Communications: Second to Third Generation Systems and Beyond*. NJ, USA : IEEE Press., 2001.

[12] L. Hanzo, F.C.A. Somerville, and J.P. Woodard, *Voice Compression and Communications: Principles and Applications for Fixed and Wireless Channels*. Chichester, UK: John Wiley-IEEE Press, 2001.

[13] S. X. Ng, J. Y. Chung and L. Hanzo, “Turbo-Detected Unequal Protection MPEG-4 Wireless Video Telephony using Trellis Coded Modulation and Space-Time Trellis Coding,” in *IEE International Conference on 3G Mobile Communication Technologies (3G 2004)*, (London, UK), 18 - 20 October 2004.

[14] E. Zehavi, “8-PSK trellis codes for a Rayleigh fading channel,” *IEEE Transactions on Communications*, vol. 40, pp. 873–883, May 1992.

[15] S. X. Ng, J. Y. Chung and L. Hanzo, “Integrated wireless multimedia turbo-transceiver design - Interpreting Shannon’s lessons in the turbo-era,” in *IEE Sparse-Graph Codes Seminar*, (The IEE, Savoy Place, London), October 2004.