EC-VIP-MC 2003, 4th EURASIP Conference focused on
Video / Image Processing and Multimedia Communications, 2-5 July 2003, Zagreb, Croatia

PERCEPTUALLY TRANSPARENT AUDIO COMPRESSION BASED ON A VARIABLE BIT RATE AAC CODER

Andrzej Szwabe[#], Czeslaw Jedrzejek[*+]

[#] Institute of Electronics and Telecommunications, Poznan University of Technology, ul. Piotrowo 3A, 60-965 Poznan, Poland, Andrzej.Szwabe@itti.com.pl

[*] Institute of Communication and Information Technologies, ul. Palacza 91A, 60-273 Poznan, Poland

[+] University of Technology and Agriculture, ul. Kaliskiego 7, 85-796 Bydgoszcz, Poland

Abstract: *The paper presents an implementation of a perceptually transparent variable bit rate (VBR) audio coder that complies with the MPEG-4 Advanced Audio Coding (AAC) standard. The coder is based on the MPEG-4 Reference Software source code of CBR coder. With small modifications in the reference source code a coder guaranteeing perceptual transparency is achieved at compression ratios considerably better than that of the reference CBR coder. Comparison tests indicate that files generated by the perceptually transparent VBR coder applied to a standard set of test items are considerably smaller (in average 24%) than the smallest CBR files of the same quality.*

Key words: *audio coding, MPEG Audio, AAC, VBR, perceptual transparency*

1. INTRODUCTION

The most popular audio coders today (including the MPEG-1 Audio Layer 3 and the MPEG-2/MPEG-4 AAC) use similar methods of a noise quantisation allocation resulting in generating bit streams of constrained variable bit rate (bit rate oscillating around a specified average value) [1, 2]. In this paper, like in MPEG documents, the term constant bit rate (CBR) will be used in reference to constrained variable bit rate [3] realised by means of noise allocation controlled by a bit reservoir. Bits are saved to the reservoir when fewer than mean_bits [2] are used to code one frame. A constant average bit rate is required for real time applications with the constant rate transmission links, but is not necessary for other kinds of applications. CBR coding can be regarded as a variable quality coding, i.e. as a mode contrasting coding at constant perceptual quality. Consequently, high quality CBR coding introduces significant bit rate overhead. This kind of overhead can be avoided in case of constant quality coding which results in a bit rate variations reflecting changing characteristics of a signal [3].

Besides a few requirements specific for variable bit rate (VBR) coding, normative MPEG-4 documents do not concern the VBR mode of AAC coding. Moreover, an informative annex of ISO/IEC 14496-3 Subpart 4 [2], describing an algorithm of a MPEG-4 AAC reference encoder, presents only a CBR mode of coding. Consequently, in a process of developing coder compliant to the standard, the normative parts were of fundamental importance as they specify requirements that concern every AAC coder. On the other hand, the description of the CBR coder provided by the informative part is a helpful source of general information on AAC encoding process and a valuable tutorial of the reference CBR algorithm.

2. A DEFINITION OF VBR AND CODING QUALITY REQUIREMENTS

The coder presented in this paper is an example of VBR coder implementation based on MPEG-4 Advanced Audio Coding (AAC) standard. AAC chosen for this implementation is commonly regarded as the state of the art in general audio coding and in many applications gradually replaces its predecessor - MPEG-1 Audio Layer 3. It has been assumed that the modified coder should be compatible with any decoder complying with MPEG-4 AAC standard.

Modifications to the reference source code were introduced with the view to achieve perceptual transparency of AAC VBR coding at compression ratios significantly better than that of the reference CBR coder. Perceptual transparency is understood as a quality guaranteeing that users sensitive to coding artefacts are not able to discriminate between an original recording of CD Audio format and a distorted one. It also has been put forward that achieving perceptual transparency at the lowest available bit rate should be possible without adjustment of any encoding parameters (in contrast to a laborious process of bit rate adjustment required in case of effective CBR coding).

3. IMPLEMENTATION

In order to achieve the best possible quality, Inner Iteration Loop of the reference coder [4] has been modified into an algorithm presented on the Fig. 1 [2].

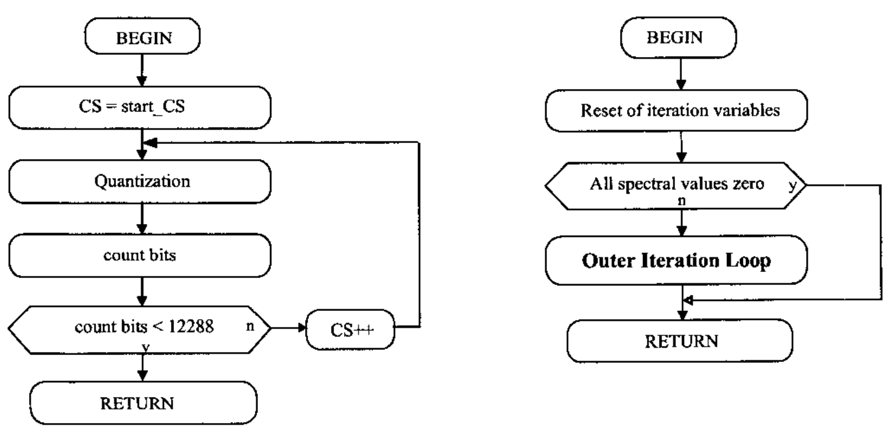

Fig. 1. The modified algorithms: Inner Iteration Loop (left) and AAC Iteration Loop (right)

When the modified Inner Iteration Loop is called, it iteratively quantises the MDCT coefficients and increases the common scale factor (CS) until the frame can be encoded with the maximum number of bits (specified by the standard). At every call of the function, the search for an optimal value of the CS begins from a start value that is specific for every frame (it is calculated by the AAC Iteration Loop so that the maximum MDCT coefficient can be encoded in the bit stream). This guarantees that the lowest possible level of the quantisation noise (represented by the CS) always corresponds to increased effectiveness of the quantisation noise equalisation performed in the Outer Iteration Loop by means of scale factors adjustment.

It should be noted that incrementation of common scale factor takes place in very rare cases when a frame initially exceeds the size of the minimum decoder input value (specified by the standard). This feature guarantees that the presented coder is always compatible with any AAC decoder. Usually the quantisation is performed using the start value of the CS.

A bit reservoir is a mechanism used to control bit rate of the CBR coder, which is useless in case of the presented VBR coder. In a consequence the AAC Iteration Loop can be simplified by exclusion of bit reservoir functions (Fig. 1).

The CBR encoder inserts FIL elements to adjust the size of raw_data_blocks upwards to the desired bit rate. Such a bit rate overhead is especially significant in case of high bit rate CBR coding. As bit rate adjusting is not required in case of VBR coding, the presented VBR coder achieves a coding efficiency gain producing bit streams that do not contain FIL elements. This advantage over CBR coding increases with a value of the highest instantaneous bit rate required to maintain perceptual transparency.

Furthermore, in case of presented VBR coder, achieving perceptual transparency at the lowest available bit rate is functionally much simpler than it is in case of CBR coding - performing encoding process is possible without an iterative search for the lowest bit rate (satisfying the perceptual transparency requirement).

4. METHODS OF CODER EVALUATION

The evaluation of the coder was a process composed of two tests. Firstly the evaluation process was carried out to verify perceptual transparency of coding. The next stage concerned evaluation of the coding efficiency achieved by the coder. Both tests were conducted using commonly used (e.g. by MPEG) high quality SQAM test items [5].

The evaluation of coding quality was performed by means of software compliant with the basic model of PEAQ method (Perceptual Evaluation of Audio Quality) as specified in ITU-R BS.1387 recommendation [6]. The obtained results have a form of Objective Difference Grade values approximating results of Subjective Difference Grade measurements and referring to a impairment scale defined by ITU-R in BS.1116 recommendation [7].

It has been assumed that results of measurements should be better than –1 (i.e. value that in ITU-R BS.1116 recommendation is referred to as "Perceptible but not annoying" distortion). It is worth mentioning that (according to an accuracy scheme specified by ITU-R working group TG 10/4) an accuracy of the PEAQ method is the highest for ODG/SDG values not worse than –1.5 [7].

In order to perform coding efficiency evaluation every test item was compressed using two coders in parallel: the presented VBR coder and a CBR reference coder. The CBR bit streams were generated at minimal bit rate resulting in the same quality as the VBR bit streams. A time-consuming process of experimental optimisation of the bit rate value can be considerably simplified using a spreadsheet simulation of a bit reservoir operation (basing on the characteristic of the VBR bit stream).

5. RESULTS OF CODER EVALUATION

The coding quality evaluation (Fig. 2) indicates that the presented coder can be regarded as approximating perceptual transparency. Besides one item (frer07_1), ODG results for all test files were better than –0.8.

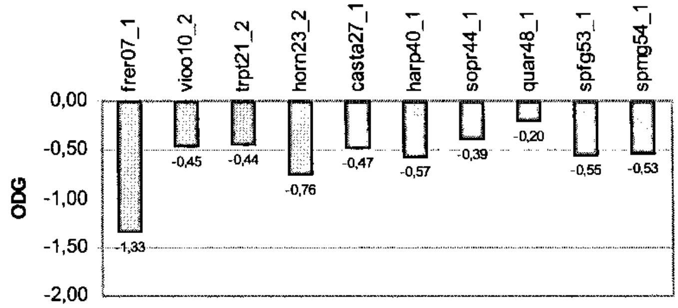

Fig. 2. The ODG quality results for SQAM test items

High quality of the encoded audio was additionally confirmed by results of informal listening tests. In case of one item (harp40_1) a distortion was hardly perceptible by one listener. In cases of the other files (including frer07_1) the coding artefacts were clearly imperceptible for all listeners.

A low ODG result in the case of an frer07_1 item may be caused by an inaccuracy of the PEAQ software. Due to a lack of transients, the test file is encoded using only long MDCT windows while the PEAQ method performs improperly in cases of distortion introduced by an extensive use of window switching [7].

The CBR reference coder operating at high bit rates closely approximates performance of the presented VBR coder. Consequently, besides very minor deviations, the results of the quality evaluation of CBR are equal to the VBR results.

On average a volume of file generated by the evaluated VBR coder was by 24 % lower than that of the CBR file. The results referring to the individual test items are presented on the Fig. 3.

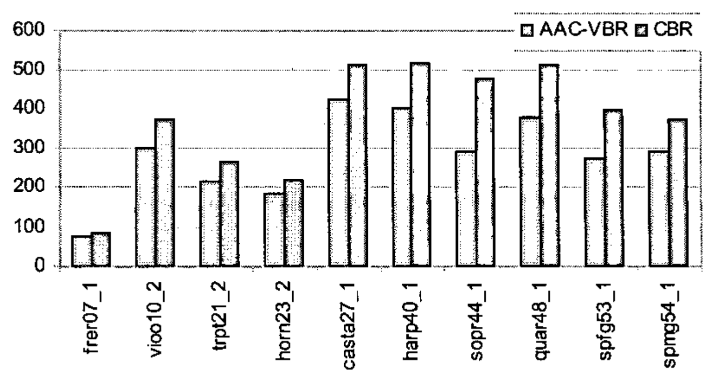

Fig. 3. Average bit rates [kbps] of perceptually transparent coding (SQAM test items) performed by the presented VBR coder as compared to bit rates achieved by the reference CBR coding (of the same quality)

In contrast to the CBR coding there is an evident dependence between spectral and temporal "complexity" of encoded audio (which can be illustrated using a spectrogram) and volume of frames generated by the VBR encoder. Comparison of consecutive VBR frames sizes to analogous sequences of frames generated by the CBR coder has been presented on the Fig. 4.

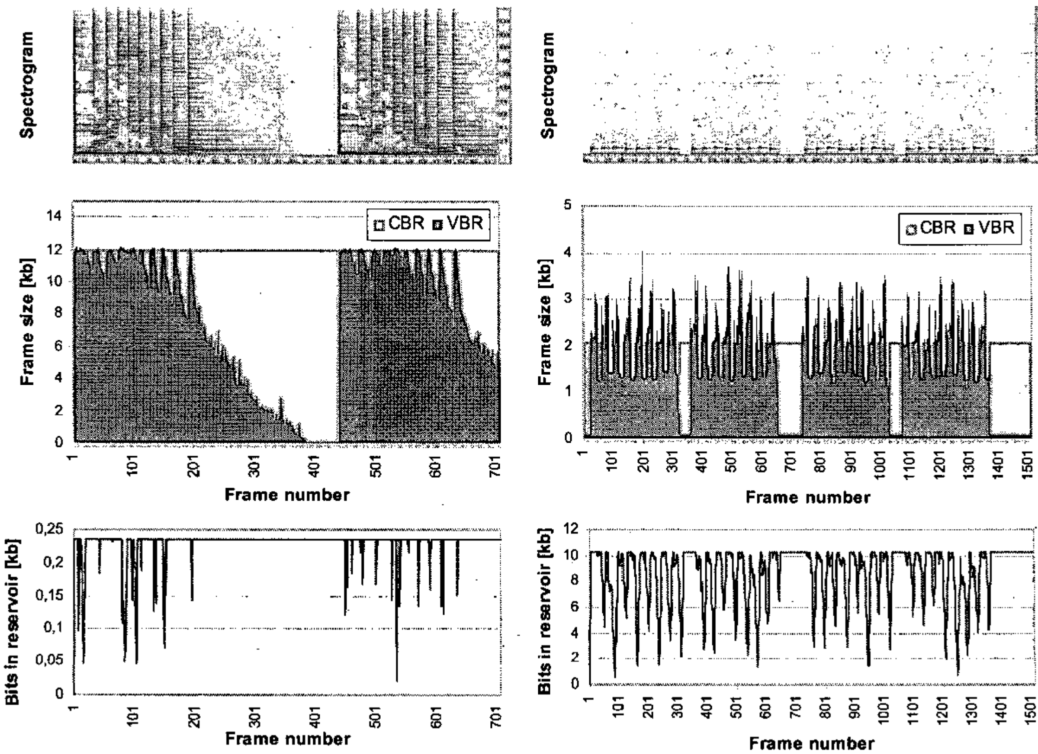

Fig. 4. Coding efficiency for two SQAM items: "spectrally complex" harp40_1 (left) and "spectrally simple" frer07_1 (right). Each item is presented by three features (corresponding to frame sequences): spectrogram (top), results of comparison of consecutive VBR frame sizes to analogous frames generated by the CBR coder, operating at the lowest bit rate resulting in a quality equal to the quality of the VBR bit streams (middle), and number of bits in a reservoir (bottom). Note that the number of bits in a reservoir cannot be negative.

As can be seen from the Figure 4, as the specified value of constrained variable bit rate decreases, larger size of bit reservoir allows higher deviations from the average value. Consequently in case of audio item that can be transparently encoded by a CBR coder operating at low bit rate, the CBR coder performance approaches the performance of the VBR coder [3].

6. CONCLUSION

The paper presents the VBR AAC coder implementation achieving perceptual transparency at compression ratio better than that of the reference CBR coder. Such a coding scheme is useful for many applications requiring high and constant quality of audio coding instead of maintaining constant bit rate of the bit stream. As it has been shown in the paper, guaranteeing such a high level of coding quality in case of CBR coding results in considerably worse compression ratio.

As for the constant quality requirement, the presented VBR coder has also an important functional feature. Achieving perceptual transparency at the lowest bit rate is possible without any adjustment of the encoding parameters. Moreover the coding efficiency considerably improves if the audio material contains fragments of silence (e.g. in case of radio speech recordings) what can simplify audio edition performed prior to encoding process.

It should be noted that the presented coder is a result of small modifications to the MPEG-4 reference source code. As the reference code has been provided by MPEG only as an example of AAC coder, the presented VBR coder (though compliant to the standard) can not be considered as a good implementation of the AAC encoder. It certainly lacks some adjustments responding to specific demands of perceptually transparent coding. For example such optimisation applied to the perceptual model would considerably improve the coding efficiency achieved by the coder. Nevertheless our work stresses importance of a AAC VBR coder (above 20% increased efficiency for the highest quality). Recently commercial implementations of a AAC VBR coder appeared on the market [8].

The AAC VBR coder is suitable in many service creation and provisioning environments. In particular it enhances flexibility of the CADENUS service architecture [9].

This work was partly supported by the IST-1999-11017 CADENUS research project. One of the authors (CJ) acknowledges the support of the BW grant from the ATR University, Bydgoszcz.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 MPEG, Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s, Part 3 Audio, International Standard 11172-3, ISO/IEC, 1993

[2] ISO/IEC JTC1/SC29/WG11 MPEG, Coding of Audiovisual Objects, Part 3: Audio, International Standard 14496-3, ISO/IEC, 1999

[3] J. Herre, Temporal Noise Shaping, Quantization And Coding Methods In Perceptual Audio Coding: A Tutorial Introduction, AES 17th International Conference on High Quality Audio Coding, 1999

[4] ISO/IEC JTC1/SC29/WG11 MPEG, Coding of Audiovisual Objects, Part 5: Simulation Software, International Standard 14496-5, ISO/IEC, 1999

[5] Sound Quality Assessment Material - Recordings For Subjective Tests, Users' handbook for the EBU - SQAM Compact Disc, Tech. 3253-E, Technical Centre of the European Broadcasting Union, 1988

[6] P. Kabal, An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality, Department of Electrical & Computer Engineering, McGill University, 2002

[7] T. Thiede, Perceptual Audio Quality Assessment using a Non-Linear Filter Bank, Ph.D. Thesis, Technical University of Berlin, 1999

[8] MPEG-4 Industry Forum, http://www.m4if.org

[9] The CADENUS Internet website, http://www.cadenus.org