FINE GRAIN SCALABLE PERCEPTUAL AND LOSSLESS AUDIO CODING BASED ON INTMDCT

Ralf Geiger[1], Jürgen Herre[2], Gerald Schuller[1], Thomas Sporer[1]

[1]Fraunhofer AEMT, Ilmenau, Germany
[2]Fraunhofer IIS-A, Erlangen, Germany
Email: ggr@emt.iis.fhg.de, hrr@iis.fhg.de, shl@emt.iis.fhg.de, spo@emt.iis.fhg.de

ABSTRACT

This papers presents an embedded fine grain scalable perceptual and lossless audio coding scheme. The enabling technology for this combined perceptual and lossless audio coding approach is the Integer Modified Discrete Cosine Transform (IntMDCT), which is an integer approximation of the MDCT based on the lifting scheme. It maintains the perfect reconstruction property and therefore enables efficient lossless coding in the frequency domain. The close approximation of the MDCT also allows to build a perceptual coding scheme based on the IntMDCT. In this paper a bitsliced arithmetic coding technique is applied to the IntMDCT values. Together with the encoded shape of the masking threshold a perceptually hierarchical bitstream is obtained, containing several stages of perceptual quality and extending to lossless operation when transmitted completely. A concept of encoding subslices is presented in order to obtain a fine adaptation to the masking threshold especially in the range of perceptually transparent quality.

1. INTRODUCTION

Scalable audio coding is a well suited technology for audio transmission over networks with dynamically changing transmission bandwidth. It provides for instance hierarchical structures with dynamically adaptable bit rates.

In MPEG-4 [1] several hierarchical scalable codecs are specified. They provide scalability between several quality levels and corresponding bitrates. The highest achievable quality level is perceptually transparent quality by using e.g. the Advanced Audio Coding (AAC) scheme.

Recently some approaches were proposed aiming combined scalable perceptual and lossless audio coding. The system proposed in [2] utilizes an MPEG-4 Audio encoder and decoder and encodes the difference (residual) signal losslessly in the time domain. In [3] this perceptual and lossless scalability is achieved by applying the Integer Modified Discrete Cosine Transform (IntMDCT). This transform allows to represent the difference values for lossless enhancement in the frequency domain and therefore can be better adapted to the underlying MDCT-based perceptual coding scheme, such as MPEG-2/4 AAC [4, 1].

This paper presents a new integrated framework for fine grain scalable perceptual and lossless audio coding based on IntMDCT. It is organized as follows: First a short review of the IntMDCT and the bitsliced arithmetic coding technique are presented. Then the concepts of perceptual significance for IntMDCT values and the coding of subslices are introduced. Finally some results for perceptual and for lossless coding are shown.

2. THE INTMDCT

The Modified Discrete Cosine Transform (MDCT) is widely used in modern perceptual audio coding schemes, such as MPEG-2/4 AAC. It provides overlapping blocks, critical sampling, and perfect reconstruction. These properties can be achieved by the concept of time domain aliasing cancellation [5]. The IntMDCT [6], [3], [7] is an integer approximation of the MDCT. It is based on the lifting scheme, introduced in [8]. This technique allows to approximate Givens Rotations by mapping integers to integers in a reversible way. To achieve this a Givens Rotation is decomposed into so-called lifting steps:

$$\begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} = \begin{pmatrix} 1 & \frac{\cos\alpha-1}{\sin\alpha} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \sin\alpha & 1 \end{pmatrix} \begin{pmatrix} 1 & \frac{\cos\alpha-1}{\sin\alpha} \\ 0 & 1 \end{pmatrix}$$

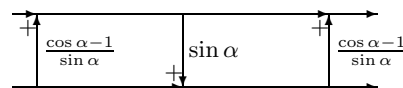Figure 1 illustrates this decomposition.

Fig. 1. Givens Rotation by three lifting steps

In every lifting step a rounding function can be included to stay in the integer domain. This rounding does not affect the perfect reconstruction property, because every lifting step can be inverted by subtracting the value that has been added.

The MDCT can be decomposed into Givens Rotations and so the lifting scheme can be applied to get an integer-to-integer approximation while maintaining the perfect reconstruction property. Figure 2 illustrates this close approximation, see also [3].

Due to the close approximation of the MDCT values a perceptual audio coding scheme can also be built upon IntMDCT instead of MDCT by applying a perceptually controlled quantization to the integer spectral values.

3. BITSLICED ARITHMETIC CODING

The concept of bitsliced arithmetic coding (BSAC) was introduced for audio coding in [10] and is standardized as a part of MPEG-4 [1]. In this context, BSAC plays the role of an alternative lossless coding kernel for MPEG-4 AAC, utilizing the MDCT and applying a perceptually controlled bandwise quantization to the spectral values. The main difference between BSAC and the standard AAC lossless coding kernel is that the quantized values are not Huffman coded, but arithmetically coded in bitslices. This allows a
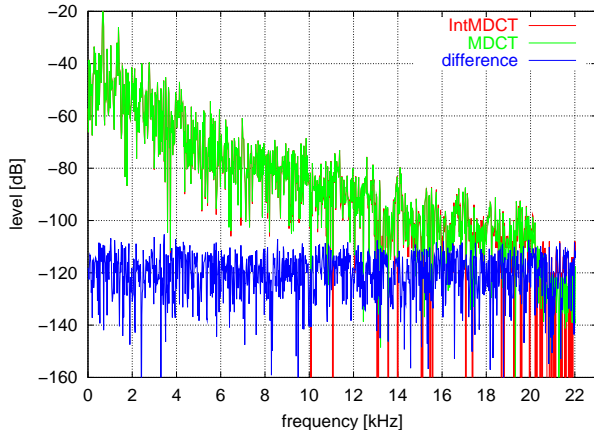
Fig. 2. IntMDCT and MDCT magnitude spectra of Carl Orff's Carmina Burana [9]

fine grain scalability by omitting some of the lower bitslices while maintaining a compression efficiency comparable to the Huffman coding approach for the quantized spectral values.

In the MPEG-4 AAC/BSAC codec the bitslices of the perceptually quantized spectral values are already ordered in a perceptual hierarchy. In this way more perceptually shaped noise is introduced as more and more bitslices are omitted. In order to adapt this coding concept to the IntMDCT values a perceptual hierarchy has to be defined. Without such a concept, omitting bitslices of the IntMDCT spectrum would merely lead to white quantization noise.

4. PERCEPTUAL SIGNIFICANCE

To achieve the least amount of audible distortion for a given number of bits, a perceptual significance is defined for the bits of the IntMDCT magnitude values, based on the permissible distortion as provided by a perceptual model. The bitslices for the hierarchical encoding are defined according to their perceptual significance. Every bitslice contains the bits with the same perceptual significance for each spectral value. This is illustrated in Figure 3.
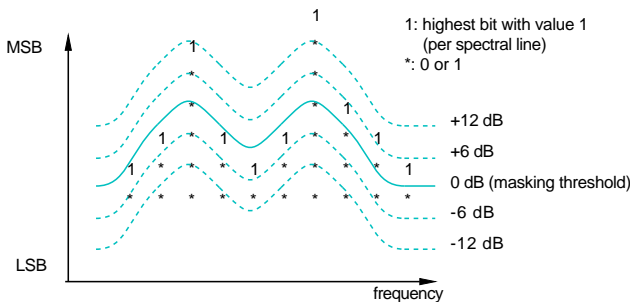

Fig. 3. Definition of bitslices with equal perceptual significance

If only some of the higher bitslices are transmitted, a quantized magnitude spectrum can be reconstructed, e.g. by simply inserting zeros for all bits that were omitted. This corresponds to a uniform quantizer with 'floor' rounding behavior. Consequently, the energy of the quantization error can be reduced by applying a

midrise quantizer with a certain offset. This is achieved by inserting a bit pattern corresponding to the desired offset (e.g. (1,0,0,...) for an offset of 0.5) for the bits that are not transmitted.

The coding of spectral coefficients' sign information can be done efficiently by transmitting the sign of a spectral value after the first non-zero bit of this spectral value is transmitted. In this way the sign value is only transmitted for spectral values that are not quantized to zero.

To reconstruct the (possibly quantized) spectral values from the hierarchically transmitted bits the decoder has to know how many bitslices are missing for a specific spectral value. In other words, the decoder has to know the perceptual significance of each spectral line. This is achieved by transmitting the masking threshold as a side information. An efficient transmission can be achieved in two alternative ways which will be discussed subsequently.

4.1. Bandwise transmission of masking threshold

The effect of masking in the frequency domain can be described along a non-linear frequency axis, i.e. the so-called bark scale. In the MPEG-4 audio coding schemes mentioned above a uniformly spaced filterbank delivers spectral bands of equal width. In order to facilitate perceptual noise shaping the spectral coefficients are grouped into frequency bands which are related to the bark scale. For each band a scale factor is determined and transmitted as a side information. The scale factor determines the quantization step-size for the spectral coefficients of the corresponding scale factor band. This concept can also be used for the efficient transmission of the masking threshold in the context of bitsliced coding of IntMDCT values.

4.2. Transmission of continuous masking threshold

The masking threshold as computed by the perceptual model in the encoder is a continuous function across frequency. Instead of using a piecewise constant function (constant within scale factor bands), it can be approximated by a polynomial, or the frequency response of a filter (LPC modeling) resulting in a closer approximation of the desired response. A 12 coefficient filter, using bark like frequency warping, is found to be generally sufficient for an approximation. These 12 coefficients are coded efficiently and transmitted to the decoder. This approach is also used for predictive perceptual coding by prefiltering [11].

5. CODING OF SUBSLICES

With the approach presented so far the accuracy of transmitted spectral values is increased by one bit with each additional bitslice, corresponding to an expected increase of the local signal-to-noise ratio by 6 dB. On the other hand, in perceptual coding it is desirable to approximate the desired precision (as determined by the perceptual model) as closely as possible in order not to "waste" bits by overcoding parts of the signal. In MPEG-2/4 AAC this leads to the adoption of finely-spaced scalefactors $2^{0.25*i}$ (with integer values i) which enable control of the quantization noise with a granularity of 1.5 dB. Thus, a stepsize of 6 dB in comparison appears too coarse to achieve both an efficient and transparent signal representation.

In order to enable a similar fine adaption of the quantization noise to the masking threshold for bitsliced coding, a subslice coding approach can be employed, as described subsequently. Each subslice contains only a few bits of one slice, and the coding of a slice is done by sequential coding of several subslices. If, for

example, a slice is divided into four subslices, every subslice only contains one out of four spectral lines. In perceptual coding the quantization noise is usually measured by the error energy in each band, each band consisting of several spectral lines. If, for example, one of the four spectral lines is refined by 6 dB by encoding an additional subslice, the total quantization error in this band is reduced by 1.5 dB on average. In this way the concept of encoding subslices allows a fine adaption of the quantization error.

Clearly, the decoder also has to know which spectral line has been enhanced in which subslice. One way to achieve this is to use a simple pattern for the spectral lines contained in each subslice. Such a pattern is illustrated in Figure 4 for the example of four subslices per bitslice. In every subslice one of four spectral lines is refined in ascending order.
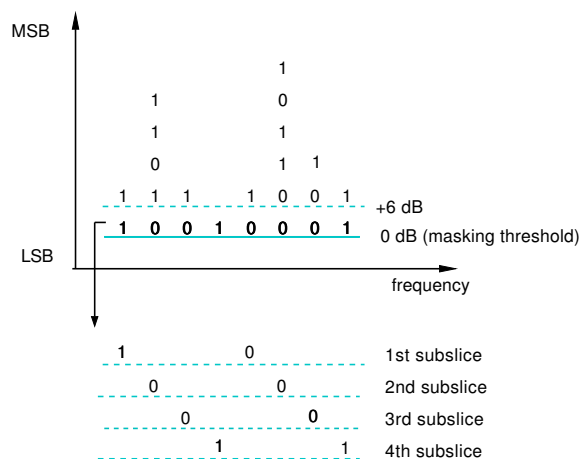

Fig. 4. Definition of subslices based on simple pattern

The simple subslice pattern approach can be enhanced by making the order of spectral line refinement in each subslice signal dependent. This can be achieved without additional side information by considering the masking threshold and the values transmitted so far in the higher subslices. In the case of the transmission of the continuous masking threshold, the assumed quantization error can be compared with the masking threshold for each line and the lines with the worst noise-to-mask ratio are refined in the next subslice. In the case of bandwise transmission of the masking threshold, all values in one band have reached the same assumed quantization error after a given bitslice is coded. Here the line order for the subslices can be based on the magnitude of the transmitted values. If the refinement in each subslice starts with the frequency bins with the smaller transmitted values, a noise shaping is achieved within each band. This is illustrated in Figure 5. In this example the '0 dB' slice has to be coded next. The order of spectral lines for the following subslices can be based on the bit values of the '+6 dB and higher' bitslices which are already transmitted. Here the spectral lines with smaller values are refined earlier than the spectral lines with bigger values.

6. RESULTS

The system was implemented with bandwise masking threshold, 4 subslices per bitslice and the noise shaping property described above. Additionally a bandwidth scaling was used. This technique is well known from the scalable audio coding schemes in MPEG-4
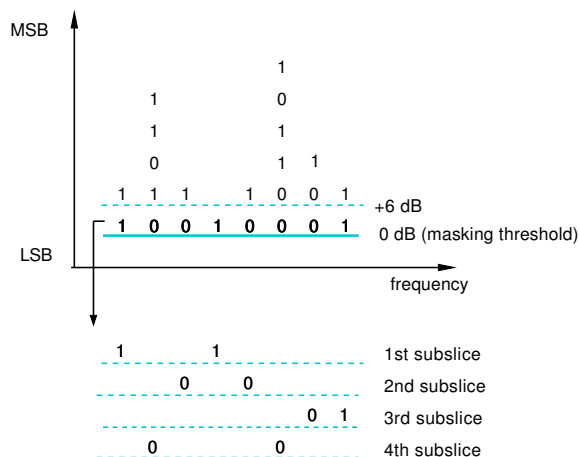

Fig. 5. Definition of subslices based on values of higher bitslices

[1] and allows a proper tradeoff between audible artifacts and audio bandwidth. In the context of bitsliced coding this is achieved by transmitting only the spectral lines of the lower frequencies in the bitslices above the masking threshold. In each subslice some high frequency lines are coded additionally to increase the bandwidth from subslice to subslice. Furthermore, this system uses the same window switching technique as MPEG-2/4 AAC. Additional coding tools such as Temporal Noise Shaping (TNS) and Mid/Side (MS) coding could also be applied to the IntMDCT spectrum (see [3]), but in this codec they are not implemented yet. As test signals the twelve critical test items used for the development of MPEG-4 Audio were coded. The input format was 44.1 kHz, stereo, 16 Bit.

6.1. Results for perceptual coding

The perceived audio quality after transmitting a certain number of subslices is evaluated based on the ITU-R BS.1387 PEAQ (Perceptual Evaluation of Audio Quality) measurement method [12] using the Noise-to-Mask Ratio (NMR) and the Objective Difference Grade (ODG) output values which are briefly described here for better understanding.

Noise-To-Mask Ratio (NMR)
The NMR estimates the ratio between the actual distortion ("Noise") and the maximum inaudible distortion, i.e. the masking threshold ("Mask"). NMR values smaller than 0dB indicate the headroom below the threshold of audibility whereas values larger than 0dB indicate audible distortions.

Objective Difference Grade (ODG)
The ODG values are designed to mimic the listening test ratings obtained from typical test listeners by means of an objective measurement procedure. The grading scale ranges from -4 ("very annoying") to 0 ("imperceptible difference").

In Table 1 the results for several quality levels are listed. The 'subslice' column represents the perceptual significance of the last subslice transmitted. In the 'bitrate' column the average stereo bitrates for these test items are listed. Currently no bitrate control is utilized, and thus the resulting average bitrate depends on the input signal (variable bitrate coding). The 'bandwidth' column lists the bandwidth chosen for the bandwidth scalability. The last two columns represent the PEAQ results. They show that the quality continuously increases when additional subslices are coded. It is, however, also visible that the worst NMR values do not improve

as fast as the perceptual significance values would indicate. One reason for this behavior could be the perceptual model used in this implementation which is currently not optimized for this application, and further improvement could be expected here. Secondly, the worst NMR values merely reflect the existence of signal portions with maximum distortion rather than the average decrease of distortion from subslice to subslice.

Subslice	Bitrate	Bandwidth	ODG	Worst NMR
+6.0 dB	60 kbps	8 kHz	-3.8	+3.0 dB
+4.5 dB	74 kbps	10 kHz	-3.8	+2.1 dB
+3.0 dB	88 kbps	12 kHz	-3.7	+2.0 dB
+1.5 dB	109 kbps	14 kHz	-3.6	+0.8 dB
0 dB	137 kbps	16 kHz	-1.8	-4.7 dB
-1.5 dB	161 kbps	18 kHz	-1.6	-6.7 dB
-3.0 dB	184 kbps	20 kHz	-1.4	-7.6 dB
-4.5 dB	205 kbps	22 kHz	-1.2	-7.5 dB
-6.0 dB	224 kbps	22 kHz	-1.0	-7.9 dB
-12 dB	305 kbps	22 kHz	-0.5	-9.3 dB
-18 dB	382 kbps	22 kHz	-0.1	-9.6 dB
-24 dB	441 kbps	22 kHz	0.0	-10.3 dB

Table 1. Average bitrates and quality results for bitsliced coding of MPEG-4 test items

6.2. Results for lossless coding

When all bitslices of the IntMDCT magnitude spectrum and all the sign values are transmitted, the signal can be reconstructed exactly in the decoder, resulting in a lossless audio coding scheme. Table 2 lists the average compression results over the signals mentioned above, and compares the results with the results for the lossless coding schemes Monkey's Audio [13] and Shorten [14]. It can be seen that the lossless compression performance of this new embedded coding scheme is comparable with the performance of other purely lossless audio coding schemes.

Lossless coder	Bitrate
Bitsliced IntMDCT	625 kbps
Monkey's Audio	569 kbps
Shorten	706 kbps
Original	1411 kbps

Table 2. Average bitrates for lossless coding of MPEG-4 test items

The good compression results for lossless coding, i.e. for coding of all bitslices, also indicate that almost no overhead is introduced by this scalable approach, and that the results for perceptual coding could be further improved by an improvement of the underlying perceptual model.

7. CONCLUSIONS

In this paper we have presented an embedded fine grain scalable perceptual and lossless audio coding scheme based on the Int-MDCT. The fine grain scalable perceptual coding is achieved by defining bitslices with equal perceptual significance and applying arithmetic coding to the bit values of the IntMDCT spectrum along these bitslices. The concept of encoding subslices allows to further refine the granularity of quantization in order to obtain a fine adaptation to the masking threshold especially in the range of perceptually transparent quality. Thus, the coding scheme allows for a large number of quality steps and corresponding bitrates for perceptual coding. Furthermore, when all the bitslices are coded, a lossless audio coding is achieved that is almost as efficient as state-of-the-art purely lossless audio coding schemes.

8. REFERENCES

[1] "Information technology - Coding of audio-visual objects - Part 3: Audio," International Standard 14496-3:2001, ISO/IEC Moving Pictures Expert Group, ISO/IEC JTC1/SC29/WG11, 2001.

[2] T. Moriya, N. Iwakami, A. Jin, and T. Mori, "A design of lossy and lossless scalable audio coding," in *Proc. ICASSP*, 2000.

[3] R. Geiger, J. Herre, J. Koller, and K. Brandenburg, "IntMDCT - A link between perceptual and lossless audio coding," in *Proc. ICASSP 2002*, Orlando, 2002.

[4] "Generic Coding of Moving Pictures and Associated Audio: Advanced Audio Coding," International Standard 13818-7, ISO/IEC Moving Pictures Expert Group, ISO/IEC JTC1/SC29/WG11, 1997.

[5] J. Princen and A. Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation," *IEEE Trans. ASSP*, vol. ASSP-34, no. 5, pp. 1153–1161, 1986.

[6] R. Geiger, T. Sporer, J. Koller, and K. Brandenburg, "Audio Coding based on Integer Transforms," in *111th AES Convention*, New York, 2001.

[7] T. Krishnan and S. Oraintara, "Fast and lossless implementation of the forward and inverse mdct computation in mpeg audio coding," in *ISCAS 2002*, Scottsdale, Arizona, May 2002.

[8] I. Daubechies and W. Sweldens, "Factoring Wavelet Transforms into Lifting Steps," Tech. Rep., Bell Laboratories, Lucent Technologies, 1996.

[9] *SQAM (Sound Quality Assessment Material)*, European Broadcasting Union (EBU), Geneva, 1988.

[10] S. Park, Y. Kim, S. Kim, and Y. Seo, "Multi-Layer Bit-Sliced Bit-Rate Scalable Audio Coding," in *103rd AES Convention*, New York, 1997, preprint 4520.

[11] G. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual Audio Coding using Adaptive Pre- and Post-Filters and Lossless Compression," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 6, pp. 379–390, September 2002.

[12] T. Thiede, W. C. Treurniet, R. Bitto, C. Schmidmer, T. Sporer, J. G. Beerens, C. Colomes, M. Keyhl, G. Stoll, K. Brandenburg, and B. Feiten, "PEAQ - The ITU Standard for Objective Measurement of Perceived Audio Quality," *Journal of the AES*, vol. 48(1/2), pp. 3–29, 2000.

[13] M. T. Ashland, "Monkey's Audio - a fast and powerful lossless audio compressor," http://www.monkeysaudio.com.

[14] A. Robinson, "SHORTEN: Simple Lossless and Near-Lossless Waveform Compression," Tech. Rep., Cambridge University Engineering Department, Cambridge, UK, 1994, Tech. Rep. 156.