STRATEGIES FOR BIT ALLOCATION REUSE IN AUDIO TRANSCODING

Mohamed F. Mansour

Texas Instruments Inc., Dallas, TX, USA
mfmansour@ti.com

ABSTRACT

We study the reuse of the bit allocation information in audio transcoding by exploiting the similarity in subband audio coding schemes. We show that important information can be deduced to reduce the encoder complexity even if the two coders employ different psychoacoustic model. We give a case study on MPEG AAC/Dolby AC-3 transcoding. The proposed algorithms can be extended to other audio transcoding schemes.

***Index Terms*—** Bit Allocation, AAC, AC-3, Transcoding.

1. INTRODUCTION

In this paper, we describe strategies for reusing the bit allocation information of the different subbands between two audio coders. We show that the bit allocation information of the first coder can be exploited to simplify the iterative bit allocation procedure in the second coder even if they use different psychoacoustic models. We study these strategies for a transcoder from MPEG-4 Advanced Audio Coding (AAC) to Dolby Digital (AC-3).

Most modern audio coding systems (including the two under study) use subband coding and psychoacoustic modeling to exploit the redundancy in the audio signal. In particular, the psychoacoustic model estimates the masking curves of each frame of the audio signal, then the encoder assigns quantization bits to different frequency bands so that the quantization noise floor is below the corresponding masking curve. If fewer bits are available, the encoder employs an optimization procedure to minimize the perceptual effect given the available number of bits. This procedure usually assigns bits to the bands with the maximum difference between the quantization noise floor and the corresponding masking curve. This basic procedure is employed with some variations in most audio coders. In this paper, we exploit this similarity to simplify the bit allocation procedure in the second encoder.

2. QUANTIZATION DISTORTION

2.1. AAC Quantization

The quantization of the AAC spectral coefficients starts by segmenting the spectrum to nonoverlapping scale factor bands. For each band, a single scale factor is transmitted for all the spectral coefficients in the band. At the encoder, the i^{th} spectral coefficient of the k^{th} scale factor band $x_{k,i}$ is scaled down by the scale factor $s(k)$ as,

$$\widetilde{x}_{k,i} = x_{k,i} * 2^{-\frac{1}{4}(s(k)-100)}$$

Then the spectral coefficients are raised to fractional power and quantized. In particular, the quantized coefficient is computed as:

$$x^{(q)}_{k,i} = Q(\widetilde{x}^{\frac{3}{4}}_{k,i}) = Q(\frac{x^{\frac{3}{4}}_{k,i}}{\Delta_k}) \qquad (1)$$

where $Q(.)$ refers to the scalar quantization function, and $\Delta_k = 2^{\frac{3}{16}(s(k)-100)}$. If we adopt the additive model for quantization noise [4], then the quantization noise random variable is defined as:

$$\delta_{k,i} = x^{(q)}_{k,i} - \frac{x^{\frac{3}{4}}_{k,i}}{\Delta_k} \qquad (2)$$

Note that $\delta_{k,i} \in [-\frac{\Delta_k}{2}, \frac{\Delta_k}{2}]$, and under some general conditions they can be approximated by an uniform independent random variables, i.e., $E\{\delta_{k,i}\} = 0$, and $E\{\delta^2_{k,i}\} = \frac{1}{12}\Delta^2_k$. At the decoder, the spectral coefficients are computed as:

$$\widehat{x}_{k,i} = x^{(q)\frac{4}{3}}_{k,i} .2^{\frac{1}{4}(s(k)-100)} \qquad (3)$$

The *overall* quantization error $\varepsilon_{k,i}$ is defined as:

$$\varepsilon_{k,i} = \widehat{x}_{k,i} - x_{k,i} \qquad (4)$$

Now, we have two cases for $\varepsilon_{k,i}$:

1. If $x^{(q)}_{k,i} = 0$, then $\varepsilon_{k,i} = \delta^{\frac{4}{3}}_{k,i}$ and

$$\begin{aligned} E\{\varepsilon_{k,i}\} &= E\{\delta^{\frac{4}{3}}_{k,i}\} = 0 \\ E\{\varepsilon^2_{k,i}\} &= E\{\delta^{\frac{8}{3}}_{k,i}\} = \frac{3}{11}(\frac{\Delta_k}{2})^{\frac{8}{3}} \end{aligned} \qquad (5)$$

2. If $x^{(q)}_{k,i} \neq 0$, then

$$\varepsilon_{k,i} = x_{k,i}(1 + \frac{\delta_{k,i}}{x^{\frac{3}{4}}_{k,i}})^{\frac{4}{3}} - x_{k,i}$$

978-1-4244-2354-5/09/$25.00 ©2009 IEEE

then by using Taylor expansion and discarding the higher order terms we get

$$\varepsilon_{k,i} \approx \frac{4}{3} x_{k,i}^{\frac{1}{4}} \delta_{k,i} + \frac{2}{9} x_{k,i}^{-\frac{1}{2}} \delta_{k,i}^{2}$$

Hence after straightforward algebra we get,

$$\begin{aligned} E\{\varepsilon_{k,i}\} &= \frac{1}{54} x_{k,i}^{-\frac{1}{2}} \Delta_k^2 \\ \sigma_{k,i}^2 &= \frac{4}{27} x_{k,i}^{\frac{1}{2}} \Delta_k^2 - \frac{1}{54^2} \Delta_k^4 / x_{k,i} \end{aligned} \qquad (6)$$

Note that, $x_{k,i}$ is not known at the decoder, however, it can be approximated in (6) by $\widehat{x}_{k,i}$. The quantization distortion cannot be estimated for frequency bands with zero scale factors. Therefore these bands are not used in the algorithm.

2.2. AC-3 Quantization

The AC-3 standard uses a different quantization procedure where each spectral coefficient has its own scale factor (called exponent component). Each spectral coefficient x_k is factored to a mantissa m_k and a 5-bit exponent e_k such that $x_k = m_k.2^{-e_k}$. The mantissa values are quantized according to the psychoacoustic model which is specified in the standard [1]. Note that $m_k \in [-1, 1]$ where this interval is segmented to equally spaced levels according to the number of bits assigned to the corresponding mantissa. If the number of bits assigned by the bit allocation algorithm to quantized n_k, then the number of quantization levels is $L_k = 2^{n_k}$. In this case, the quantization error $\varepsilon_k \in [-\frac{2^{-e_k}}{L_k}, \frac{2^{-e_k}}{L_k}]$ and the variance of the quantization noise is:

$$E(\varepsilon_k^2) = \frac{4^{-e_k}}{3L_k^2} \qquad (7)$$

Note that, the use of the number of quantization levels in the above equation is more adequate than using the number of assigned bits as the number of quantization levels in the AC-3 standard is not always power of two. In this case, we can directly substitute L_k from the standard quantization tables.

3. REUSE ALGORITHM

3.1. AC-3 bit allocation algorithm

We start our discussion with a brief overview of the AC-3 bit allocation algorithm as described in the standard [1]. The psychoacoustic model is part of the AC-3 standard in contrast to MPEG standards which do not mandate the encoder. Therefore the bit allocation algorithm is completely specified in the AC-3 standard [1]. The objective of the reuse algorithm is to reduce the number of iterations required in this procedure by exploiting the bit allocation information in the AAC bitstream. In the following, we briefly overview the main components of the AC-3 bit allocation algorithm.

The objective of the bit allocation algorithm is to decide the number of bits assigned to the mantissa of each spectral coefficient. It employs a parametric model of the human hearing. This parametric model is computed at the decoder to generate the bit allocation information, i.e., the detailed bit allocation information is not included in the bitstream, rather, only minimal side information is sent. The parametric model uses a set of excitation functions that represent the offset in the amplitude that can be tolerated in the quantization process. These excitation functions are derived from the exponent component of the spectral coefficients[3]. The masking curves are derived as the minimum of the excitation functions and the absolute hearing threshold at the different frequencies. These masking curves are modified by coarse and fine offset values for each channel (referred to as "*csnroffset*" and "*fsnroffset*" respectively in [1]). The difference between the modified masking threshold and the actual spectral value is then used to decide the number of bits assigned to the corresponding coefficients using lookup tables. At the encoder, all the previous steps are done only once, then the encoder iterates on the offset values to find the optimal value that utilizes the largest number of bits without exceeding the bits pool size. If for a given iteration, the current of values of "*csnroffset*" and "*fsnroffset* requires a number of quantization bits that exceed the pool, then their values are reduced in the following iteration and vice versa until reaching their maximum value within pool size. The result of the bit allocation algorithm are the parameters '*csnroffset*" and "*fsnroffset* which are included in the bitstream for each frame. For more detailed description of the bit allocation algorithm, refer to [1],[3].

3.2. Proposed algorithm

The basic idea of the reuse algorithm is to match the quantization distortions in the corresponding frequency bands in both AAC and AC-3 coders after compensating for the filter delay in the AAC synthesis filter bank and the AC-3 analysis filter bank. Exact matching of the distortion is not expected due to the difference in the psychoacoustic model and the number of channels. Rather, we derive bounds on the AC-3 distortion that are derived from the corresponding distortion in the AAC data. These bounds are used to limit the search space of "*snroffset*" parameter in the AC-3 bit allocation algorithm resulting in reducing the number of iterations.

The first step of the algorithm is to choose the frequency bands for comparison. A small fraction of bands is used for matching purposes. Since the AC-3 coder has 256 channels and AAC coder has 1024 channels, then roughly for stationary signals each AC-3 band corresponds to four AAC bands. The time resolution of the AC-3 frame is higher than the AAC time resolution, however, for stationary parts of the signal we use the inherent assumption of subband coding schemes that the spectral characteristics do not change over the span of a whole frame. Therefore the optimized bit allocation al-

gorithm is used only when both the AAC and the AC-3 coders use long blocks for the corresponding frames. The standard AC-3 bit allocation algorithm is used in case of short blocks in either coder, where the bands mapping becomes rather complicated. Note that the long blocks account for more than 90% of all frames in most audio signal.
The matching frequency bands are usually in the lower side of the spectrum where typically most of the energy is concentrated. However, the few bands next to DC are not used to mitigate the effect of high pass filtering that is usually employed in the encoder to enhance the signal perception. The typical number of the matching AC-3 bands is four bands (which correspond to 16 AAC bands) in the range of bands between 10-40. An adaptive algorithm for selecting the matching bands is described in section 3.3.1.
Assume that the matching AC-3 frequency bands are between N_1 and N_2 (i.e., the corresponding AAC bands are $4N_1$ and $4N_2$). Define a scaling factor λ that scales the AAC distortion to the AC-3 distortion (where λ is a function of the bit rates of both the AAC and AC-3, and it is computed offline using training sequences). The optimized bit allocation algorithm proceeds as follows:

1. Compute the AAC distortion of the bands between $4N_1$ and $4N_2$ according to (5) and (6). Compute the maximum and minimum distortions d_{max} and d_{min}.

2. Run the AC-3 bit allocation algorithm for the bands between N_1 and N_2. At each iteration, compute the average distortion of these bands according to (7). If the distortion is higher than λd_{max} then increase $snroffset$ parameters and vice versa until convergence. Denote the final $snroffset$ value by $off1$. Note that the computational complexity of this step is small as the bit allocation algorithm is run over a small number of bands (typically 4 bands) as opposed to 256 bands of the full bit allocation algorithm.

3. repeat the previous step for λd_{min} to compute $off2$.

4. Run the full AC-3 bit allocation algorithm with $off1$ and $off2$ as upper and lower bounds on $snroffset$ value.

5. The above steps are performed only when both AAC and AC-3 coders use long window blocks. If either of them uses short window blocks then the standard bit allocation algorithm is used instead.

Note that, we did not explicitly incorporate the psychoacoustic model of the first coder. However, it is inherently reflected in the quantization step of the spectral coefficients. The overhead of the above algorithm includes the computation of the quantization distortion in both AAC and AC-3 coders. This is done using lookup tables on a small fraction of coefficients which adds small computational complexity. The algorithm significantly reduces the search span of $snroffset$ values, therefore it reduces the number of iterations before convergence.
The basic procedure of the algorithm is intuitive in principle, however, many implementation issues are addressed in the following section that render it practically useful for a wide variety of audio signals.

3.3. Implementation Issues

3.3.1. Choosing Bands

The matching bands can be chosen adaptively to cope with the signal dynamics. Rather than using a fixed set of bands, the AAC bands with highest energy at each channel are used. This improves the estimation of the quantization noise variance and reduces the possibility of having bands with zero scale factor. The bands with highest energy are computed from the scale factor data of the AAC bit stream. The search is started in the bands after the DC to avoid the possible effect of the high pass filtering preprocessing at the encoder. The selected bands are the bands with highest minimum scale factor.

3.3.2. Joint Stereo Procedure

Joint stereo coding is usually employed in the AAC coder to reduce the bitrate of the stereo signals by encoding the sum and difference of the stereo signals. The rematrixing procedure in the AC-3 encoder resembles the AAC joint stereo coding. If both the AAC and AC-3 employ joint stereo coding, then the reuse algorithm is used as before as both channels are reformed in a similar way. However, if one coder uses joint stereo coding while the other does not, then the reference bit allocation algorithm is used instead.

3.3.3. Temporal noise Shaping (TNS)

TNS employs linear prediction in the spectral domain to shape the quantization noise in the time domain to match the signal energy. The estimation of the AAC quantization noise variance involves IIR filtering if the TNS exists. This significantly increases the complexity of the noise estimation. Therefore frequency bands that use TNS are discarded from the search procedure described in section 3.3.1. This is not usually a concern as TNS is usually employed in mid-frequency bands and the lowest bands are quantized directly.

3.3.4. Short Windows

In principle, the proposed algorithm can be used even if the coders use short windows. However the inherent stationarity assumption is violated in this case. Since the AC-3 bit allocation procedure is global for a whole frame of 1536 samples, then mapping from the AAC data during transients will not

File	Genre	Duration (sec)	Iterations
Castanets	Percussive	6.53	10.5%
Guitar	Single Instrument	3.82	24.2%
Bach	Classic	24.94	15.4%
Funky	Pop	19.71	13.7%
Spot	Pop	10.73	14.5%

Table 1. Complexity Comparisons of the Proposed algorithm

be accurate. Therefore the standard AC-3 bit allocation algorithm is used in this case. However, this is not a concern as for most audio signals the short blocks typically account for less than 5% of the total frames.

4. EVALUATION AND DISCUSSION

We evaluated the proposed algorithm for many test audio signals. The evaluation measure of the proposed algorithm is the total number of iterations needed for the convergence of the bit allocation algorithm to its final value. Note that, the final convergence value is the same for both the optimized and standard bit allocation algorithm. The improvement is in the number of iterations needed to reach this point. In table 1, we listed the reduction in the total number of loops when the AAC is running at 128 kbps and the AC-3 is at 192 kbps.

From the table we notice that the total number of iterations of the bit allocation algorithm is significantly reduced in most cases. The improvement is more apparent when the test signal has more spectral variation and does not obey the average parameters that are used in the standard bit allocation algorithm.

Another evaluation measure is the perceptual distortion when a fixed maximum number of bit allocation iterations is used. To measure the perceptual distortion, we used the EAQUAL software [6] which is an implementation of an ITU-T standard for evaluating high quality audio coders. We used the Objective Difference Grade (ODG), which compares the decoded file to a reference file and uses a quality rating that is based on ITU-R BS.1116 guidelines [5]. The ODG values are in the range [-4,0] where -4 stands for very annoying and 0 stands for imperceptible difference between the reference and test files. Although the absolute value of the ODG is not very accurate, the relative difference represents a good indication of the audio quality improvement. In table 2 we list the quality tests when the AAC operates at 256 kbps and the AC-3 at 192 kbps. We used a high bit rate for the AAC decoder to minimize the degradation introduced by the first coder. We used only three iterations for both the reference bit allocation algorithm and the optimized bit allocation algorithm. The last column in the table represents the final ODG value when infinite number of iterations are allowed till convergence.

File	ODG (Reference Alg.)	ODG (Proposed Alg.)	ODG(∞)
Castanets	-1.24	-0.71	-0.61
Guitar	-1.09	-1.03	-1.01
Bach	-2.08	-1.90	-1.56
Funky	-1.12	-0.83	-0.60
Spot	-0.99	-0.83	-0.64

Table 2. Perceptual Performance Comparison

We note from the table that there is a noticeable difference in the quality on the ODG scale, and the proposed algorithm converges much faster to the final value.

Note that, in this paper we considered two coders that employ different psychoacoustic models. We were able to achieve noticeable reduction in the bit allocation complexity. If both coders employ a similar psychoacoustic model and operate at the same bit rate, then the bit allocation algorithm can be further optimized. For example, instead of computing the quantization distortion, it is expected that both coders will have the same number of zero bands (where no bits are assigned to the whole band). Therefore, the last nonzero band in the two coders can be matched instead of computing the quantization distortion.

In summary, we presented a framework for reusing the bit allocation information in audio transcoding schemes based on matching the quantization distortion of both coders. We showed that the proposed algorithm provided significant reduction in the bit allocation complexity of the encoder even if different psychoacoustic models are employed.

5. REFERENCES

[1] "Digital Audio Compression Standard (AC-3, E-AC-3,) Revision B", Document A/52B, Advanced Television Systems Committee, 2005.

[2] ISO/IEC 14496-3, Information technology – Coding of audio-visual objects – Part 3: Audio, 1999.

[3] C. Todd; G. Davidson; F. Davis; L. Fielder; D. Link; S. Vernon,"Parametric Bit Allocation in a Perceptual Audio Coder", 96th AES convention, Feb. 1994.

[4] R.Gray and D. Neuhoff,"Quantization", IEEE Trans. Inform. Theory, vol. 44, pp. 2325-2383, Oct. 1998.

[5] "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems", International Telecommunications Union, Reccommendation ITU-R 1116, 1994.

[6] A. Lerch, EAQUAL Evaluation of Audio Quality: http://www.mp3-tech.org/programmer/sources/eaqual.tgz.