SCALABLE AUDIO CODING USING WATERMARKING

Mahmood Movassagh *Peter Kabal*

Department of Electrical and Computer Engineering
McGill University, Montreal, Canada
Email: {mahmood.movassagh@mail.mcgill.ca, peter.kabal@mcgill.ca}

ABSTRACT

A scalable audio coding method is proposed using a technique, Quantization Index Modulation, borrowed from watermarking. Some of the information of each layer output is embedded (watermarked) in the previous layer. This approach leads to a saving in bitrate while keeping the distortion almost unchanged. This makes the scalable coding system more efficient in terms of Rate-Distortion. The results show that the proposed method outperforms the scalable audio coding based on reconstruction error quantization which is used in practical systems such as MPEG-4 AAC.

Index Terms— Scalable coding, Quantization Index Modulation, Watermarking, Entropy coding.

1. INTRODUCTION

Bit-rate scalability has been a desired feature in multimedia communications. Without the need to re-encode the original signal, it allows for improving the quality of an audio/video signal as more of a total bit stream becomes available, or lowering the quality if channel conditions deteriorate. Scalability can also provide robustness to packet loss for transmission over packet networks. In such systems, robust channel coding can be performed for the core bitstream so that all the receivers can receive it without loss. The rest of the bitstream is sent with normal channel coding. Thus, if the enhancement layers are lost, the signal can be still reconstructed at base level quality.

Several scalable coding systems have been proposed so far, including using wavelet transforms [1], bit-plane based coding [2, 3], and fine-grain scalable coding [4, 5]. One popular scalable coding system, at the core of AAC scalable coding [6], is a system based on reconstruction error quantization (REQ). In REQ (Fig. 1), the signal is quantized by an quantizer designed for a minimum bit rate and acceptable distortion (the base layer). Enhancement layers improve the quality of the base layer signal, refining the quantization by subtracting the quantized signal from the original. This error signal is quantized, encoded and transmitted as the first enhancement layer. This enhancement step can be repeated, to form an ordered set of layers. From the base up, each additional layer that the receiver receives is used to refine the quality of the decoded signal.

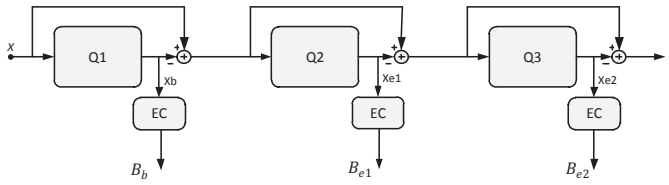

Fig. 1. Salable Audio Coding based on REQ. EC is entropy coding.

In terms of Rate-Distortion (RD) performance, REQ is optimal for the Mean Square Error (MSE) criterion. It asymptotically achieves the performance of an equivalent non-scalable coding system [7] if the rate is measured by the entropy of resulting output symbols. However, in practical coding systems symbols need to be encoded in a bitstream using an entropy coding scheme, which adds an overhead for each layer.

In this paper we propose a method that improves the performance of the scalable audio coding systems base on REQ. In this method we used the Quantization Index Modulation (QIM) which is a technique borrowed form watermarking [8]. Using this technique some of the information of each layer output is embedded (watermarked) in the previous layer. We will show that using this approach, a saving in bit rate is achieved while it does not much affect the distortion.

In the following section we will discuss AAC quantization and will address an issue which we will take advantage of in our method to improve the performance of the REQ system. Then in Section 3 we will give a brief introduction to QIM technique and in Section 4 we will introduce our proposed method Scalable Audio Coding using Watermarking which we will refer to by WSAC in the rest of paper. Simulation results will be presented in Section 5 and the paper will conclude in Section 6.

2. AAC AND UNIFORM THRESHOLD QUANTIZATION

The quantization formula which is used in AAC [9] is given by

$$\begin{aligned} i_x &= \text{sgn(x)nint}(\frac{|x|^{0.75}}{\Delta} - 0.0946) \\ \hat{x} &= \text{sgn(x)}(\Delta|i_x|)^{\frac{4}{3}}, \end{aligned} \tag{1}$$

where Δ is the step size parameter, nint() and sgn() denote the nearest integer and signum functions and 0.0946 is the offset value which is also referred to as the magic number. This quantization formula is based on the assumption that the input signal is Laplacian. The optimal quantizer for exponential and Laplacian signals, a special case of exponential signals, is a Uniform Threshold Quantizer (UTQ) with a dead-zone around zero [10]. In such a quantizer, reconstruction points are not in the middle of the quantizer intervals and there is a constant offset in each interval. However, the interval width (or the step size) of the quantizer is constant (uniform threshold) except for the zero interval (the dead-zone) which is larger than the other intervals.

In high-rate theory, for the case of constrained-entropy quantization the relation between the optimal uniform quantizer step size and its entropy is [11]

$$\Delta = 2^{-(R-h(X))}, \tag{2}$$

where Δ is the quantizer step size, R is the quantizer output entropy and $h(X)$ is the input signal differential entropy. This equation implies that when you double the resolution of the quantizer (by halving the step size) the output entropy will be increased by one bit. The above relation holds only with the high-rate assumptions where 1) quantization cells are small enough that the source density is constant within the cell 2) each re-construction point is located at the center of the cell and 3) $N \to \infty$. However, in a REQ scalable coding system a low resolution quantizer is used in each layer, e.g. a 4-layer system with 4-bit quantizers. In the following we will obtain a general relation for the entropy of a UTQ and will show that the entropy increment becomes less than unity for the low-rate quantization.

Consider a Laplacian signal which is a good approximation for audio signals. Note that this is true even for the compressed MDCT coefficients. In AAC the compression is performed on the input signal MDCT coefficients before quantization ($x^{0.75}$ function in the above formula). In the rest of the paper, by input signal we mean the MDCT coefficients after this compression. For a Laplacian signal the pdf is

$$f_X(x) = \frac{1}{2}\lambda e^{-\lambda|x|}, \tag{3}$$

where $\lambda = \frac{\sqrt{2}}{\sigma}$ and σ is the standard deviation of the signal.

Now consider a UTQ. The probability of the input signal to be in each interval (for $i > 0$) is

$$p_i = \frac{1}{2}(e^{-\lambda t_i} - e^{-\lambda t_{i+1}}), \tag{4}$$

where t_i and t_{i+1} are the thresholds of the interval i and we have (except the dead-zone)

$$t_{i+1} = t_i + \Delta \tag{5}$$

which gives

$$p_i = \frac{1}{2}(1 - e^{-\lambda\Delta})e^{-\lambda t_i} \tag{6}$$

and the probability of the dead-zone becomes

$$p_0 = (1 - e^{-\lambda T}), \tag{7}$$

where $T = \frac{\Delta}{2} + t_{offset}$ (t_{offset} is the offset value of the UTQ).

The entropy of the quantizer output then can be written as

$$\begin{aligned} R &= -\sum_i p_i \log_2(p_i) = -(1 - e^{-\lambda T}) \log_2(1 - e^{-\lambda T}) \\ &\quad - 2\sum_{i\geq 1} \frac{1}{2}(1 - e^{-\lambda\Delta})e^{-\lambda t_i} \log_2(\frac{1}{2}(1 - e^{-\lambda\Delta})e^{-\lambda t_i}) \\ &= (1 - e^{-\lambda T}) - (1 - e^{-\lambda\Delta})[\sum_{i\geq 1} e^{-\lambda t_i} \log_2(e^{-\lambda t_i}) + \\ &\quad (\log_2(1 - e^{-\lambda\Delta}) - 1)\sum_{i\geq 1} e^{-\lambda t_i}]. \end{aligned} \tag{8}$$

Since $t_1 = T$, we have

$$\begin{aligned} &\sum_{i\geq 1} e^{-\lambda t_i} \log_2(e^{-\lambda t_i}) = e^{-\lambda T} \log_2(e^{-\lambda T}) \\ &\quad + e^{-\lambda(T+\Delta)} \log_2(e^{-\lambda(T+\Delta)}) \\ &\quad + e^{-\lambda(T+2\Delta)} \log_2(e^{-\lambda(T+2\Delta)})) \\ &\quad + \dots \\ &= e^{-\lambda T} \log_2(e^{-\lambda T})(1 + e^{-\lambda\Delta} + e^{-2\lambda\Delta} + \dots) \\ &\quad - \lambda\Delta \log_2(e) e^{-\lambda T}(e^{-\lambda\Delta} + 2e^{-2\lambda\Delta} + \dots) \\ &= \frac{e^{-\lambda T} \log_2(e^{-\lambda T})}{1 - e^{-\lambda\Delta}} - \frac{\lambda\Delta \log_2(e) e^{-\lambda T} e^{-\lambda\Delta}}{(1 - e^{-\lambda\Delta})^2} \end{aligned} \tag{9}$$

and

$$\sum_{i\geq 1} e^{-\lambda t_i} = e^{-\lambda T}(e^{-\lambda\Delta} + e^{-2\lambda\Delta} + \dots) = \frac{e^{-\lambda T}}{(1 - e^{-\lambda\Delta})}, \tag{10}$$

then

$$\begin{aligned} R = e^{-\lambda T} &+ \lambda \log_2(e) e^{-\lambda T}(T + \frac{\Delta e^{-\lambda\Delta}}{1 - e^{-\lambda\Delta}}) \\ &- e^{-\lambda T} \log_2(1 - e^{-\lambda\Delta}) \\ &- (1 - e^{-\lambda T}) \log_2(1 - e^{-\lambda T}). \end{aligned} \tag{11}$$

The quantization loading factor L is defined as $L = x_m/\sigma$, where σ is the standard deviation of the input signal and x_m is the quantization limit which is for instance $2^{13} - 1$ in AAC quantization. By properly choosing this loading factor, the probability of signal to be beyond the quantization limit becomes negligible. That is why we used the infinite summation formula for the above geometric series (in all resolutions the quantizer limit is fixed and that probability remains the same for a given L). Normally $L \geq 7$ is chosen for Laplacian signals to avoid the quantization overloading effect. In Fig. 2 the obtained entropy versus the quantization resolution was plotted for the AAC quantizer and for three different values of L. Note that changing the step size parameter Δ in (1) changes the quantizer resolution at the same time. For instance, if $\Delta = 2$ then the resolution of the quantizer is halved. In other words the resolution (number of levels) can be expressed by: $N = 2x_m/\Delta$. The horizontal axis in Fig. 2 is $\log_2(N)$.

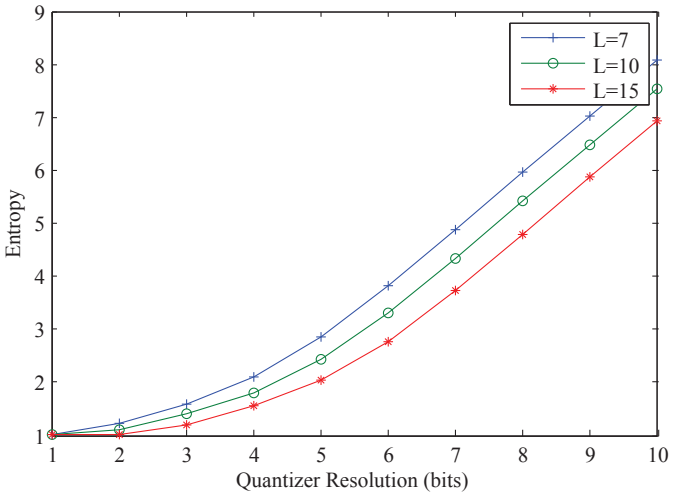

Fig. 2. Entropy of a UTQ versus quantization resolution for three values of input σ ($L = 7$, $L = 10$ and $L = 15$)

It can be seen that for high resolutions the slope of the curves $\frac{\Delta R}{\Delta b}$ tends to unity (as predicted by the high-rate equatio), while for lower resolutions it is less than unity. This fact that in low resolutions the entropy increment becomes less than unity for doubling the resolution is the issue that we mentioned in the introduction section. We will take the advantage of that in our proposed coding system.

3. QUANTIZATION INDEX MODULATION (QIM)

QIM is one of the techniques used in watermarking. Consider for instance the AAC quantizer in which the quantized values are integers (nint() function in (1)). The idea is to quantize a signal to the nearest-even or nearest-odd integer rather than to the nearest integer value. If the receiver receives an even value a 0 bit is decoded and otherwise a 1 bit is decoded. Therefore, QIM enables us to watermark one bit of data into the signal at the cost of higher distortion. Figure 3 shows the QIM based on the nearest-even integer quantization.

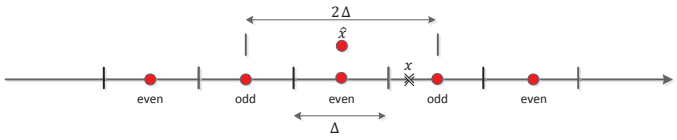

Fig. 3. Quantization Index Modulation and its equivalency to quantization with half resolution

However, the point (which is desired here) is that if we double the resolution of a quantizer and use the QIM, that is equivalent to using the original quantizer in terms of rate-distortion. Figure 4 shows the rate distortion performance of an AAC quantizer without and with using QIM denoted by UTQ and QIM respectively ($L = 7$). For each resolution, first the regular quantization was performed using UTQ, and then the resolution was doubled and QIM was applied to the new quantizer for a randomly generated input bitstream with $0.5 - 0.5$ probabilities.

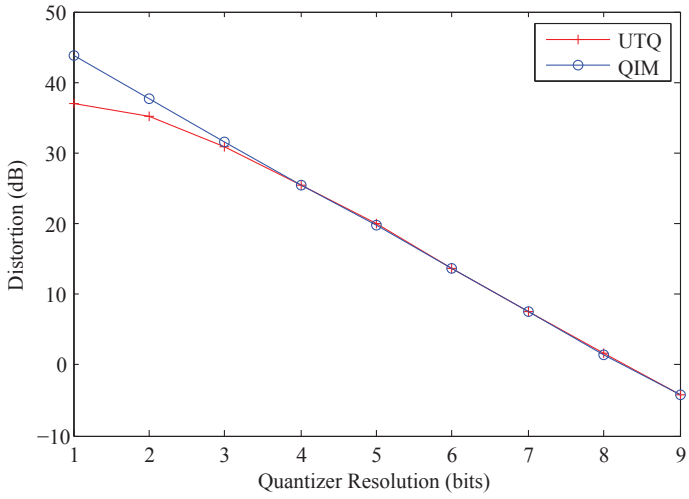

Fig. 4. Performance of UTQ and its equivalent QIM ($L = 7$)

As can be seen in the figure, the difference in performance of UTQ and its equivalent QIM is not noticeable for $b \geq 3$. Note that in performing QIM using the nearest-even integer function (NE-QIM) the dead-zone covers one interval, whereas in the nearest-odd integer QIM (NO-QIM) two adjacent intervals around zero are covered in the dead-zone. However, as can be seen in Fig. 4, that does not much affect the performance for the desired range of resolution ($b \geq 3$). Equation (11) gives the entropy of a UTQ in terms of the step size and offset values. When using the NE-QIM approach,

the same equation can be used for obtaining the entropy since the two quantizers are exactly equivalent. However, when using the NO-QIM since there are two dead-zones around zero the entropy equation for such a quantizer slightly differs from (11) only in the last term. Nevertheless, the same story for the differences in entropy is repeated. In fact, the weighted average entropy of the two quantizers (in NE-QIM and NO-QIM) also follow the same pattern of Fig. 2 if plotted.

4. SCALABLE AUDIO CODING USING WATERMARKING

In a scalable audio coding based on REQ, after quantization in each layer entropy coding is performed. The entropy coding used in AAC for this scalable coder is Huffman coding which we use here too. For each quantization resolution a Huffman codebook is built which is used by both the coder and the decoder. In these codebooks there are codewords corresponding to each quantizer reconstruction point which is found in the entropy coding process and sent to the receiver. These codewords are variable length and their average determines the bitrate of each layer.

The main idea in our proposed method is that for each layer output we embed one bit of each codeword in the previous layer using QIM. By this approach the average bitrate of each layer is exactly decreased by unity (except the base layer which has no preceding layer):

$$\sum_i p_i(w_i - 1) = \sum_i p_i w_i - \sum_i p_i = B - 1, \qquad (12)$$

where B is the average bitrate and w_i is are the codeword lengths.

On the other hand, since we are using QIM for the previous layer, to keep the distortion unchanged we need to double the quantization resolution of that layer. This will be done for all layers except the last one for which QIM is not performed (it has no succeeding layer). However, as we showed in section 2, in low resolution (which is the case for each layer of REQ scalable coding), when we double the quantization resolution its output entropy (hence the bitrate after entropy coding) is increased by less than one. Therefore, we save one bit in bitrate of a layer by watermarking in the previous layer, and at the same increase the bitrate of that layer by an amount less than one by doubling the quantization resolution which in total leads to a savings in bitrate:

$$\begin{aligned} B' &= (B_1 + \Delta B_1) + (B_2 + \Delta B_2 - 1) + (B_3 + \Delta B_3 - 1) \\ &\quad + \cdots + (B_M - 1) \\ &= B - [(M-1) - \Delta B_1 - \Delta B_2 - \cdots - \Delta B_{M-1}], \end{aligned} \qquad (13)$$

where B' is the new total bitrate using WSAC, B is the total bitrate using REQ, B_i is the bitrate of layer i before using QIM, M is number of layers and ΔB_i is the increment in its bitrate after using QIM which is less than unity. Assuming that the same resolution is used for all layers and the bitrate increment for them equals to ΔB we can write

$$B' = B - (M-1)(1 - \Delta B). \qquad (14)$$

This gives a clear equation for savings in bitrate which equals to $(M-1)(1-\Delta B)$. The more number of layers, the more we save in bitrate.

4.1. Coding

Consider the REQ scalable coding system (Fig. 1). First we perform coding based on REQ from first layer to the last. However, instead of using regular quantizers we use doubled-resolution quantizers and in each layer apply NE-QIM to them. In the next step, from the last layer to the first we perform the following algorithm:
If the ith bit of the output codeword for the current layer is '1', requantize the input of the previous layer using NO-QIM and obtain the new codeword. Otherwise, keep the codeword obtained in the first step.

This algorithm is repeated for each layer. Note that the choice of i is made based on the resulting probabilities of the watermarked bits which directly affects the performance of the system. A specific i should be determined and be known by both coder and decoder. Since the codewords are variable length and some of them might have a length shorter than i, circular shifting can be used to solve this problem. Using this technique in our simulations shows that considering the last bit of codewords for watermarking leads to good results.

Figure 5 shows the block diagram of WSAC for a 3-layer WSAC. Note that the last layer remains unchanged since it does not have any succeeding layer.

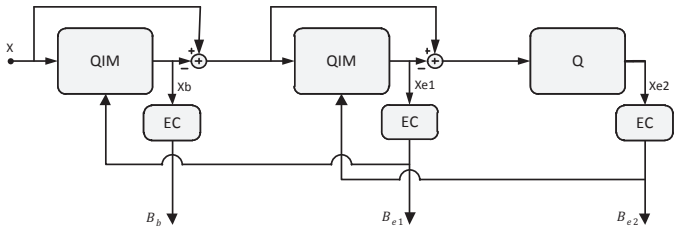

Fig. 5. Block diagram of a 3-layer WSAC

4.2. Decoding

The receiver starts decoding from the base (first) layer to any level it receives. For each layer, after decoding the variable length codeword, if the quantization index is even the ith bit of the next layer output codeword is considered to be '0'. Otherwise it is '1'. This bit will be used as the ith bit of the next layer's output codeword (if received).

In the REQ system. We simply add the outputs of the layers to reconstruct the original signal in a scalable manner. Here, however, a modification is required for the reconstruction pattern. Consider a 2-layer WSAC. In the first step we use NE-QIM for the first layer. In the next step, however the NO-QIM might be used for this layer depending on the second layer output. Reconstruction of the original signal at the receiver is dependent on which QIM is used for the first layer since the input of the second layer, which is the reconstruction error of the first layer, was formed in the first step of QIM in which NE-QIM was used for the layers. See Fig. 6. The reconstruction error resulting from NE-QIM is denoted by e_e in the figure. If the decoded quantizer index for the first layer is odd, which means NO-QIM was applied to the first layer, for reconstructing the original signal there are two possible cases: e_e is positive or negative. If e_e is positive the reconstruction formula is

$$x = \hat{x} + e_e - \Delta, \tag{15}$$

and for the negative case

$$x = \hat{x} + e_e + \Delta. \tag{16}$$

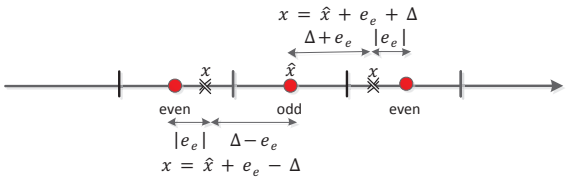

Fig. 6. Reconstruction scenarios in WSAC

If the decoded quantizer index for the first layer is even, reconstruction is simply performed by

$$x = \hat{x} + e_e. \tag{17}$$

Note that what receiver actually uses is the quantized version of e_e which is the output of the next layer.

This is the reconstruction algorithm used for all layers in WSAC. Note that for each layer the appropriate step size Δ should be used which might be different from other layers.

5. SIMULATION AND RESULTS

Two common REQ cases were considered for our simulation. A 4-layer system with 4-bit quantizers in each layer and a 5-layer system with 3-bit quantizers. For WSAC, since the resolutions are doubled the corresponding quantizers used were 5-bit and 4-bit respectively. The quantizations were performed using AAC quantization formula. A set of Huffman coding tables were generated for different resolutions which were used for entropy coding in both REQ and WSAC. Laplacian random variables were generated as the input signal with the desired values of L parameter.

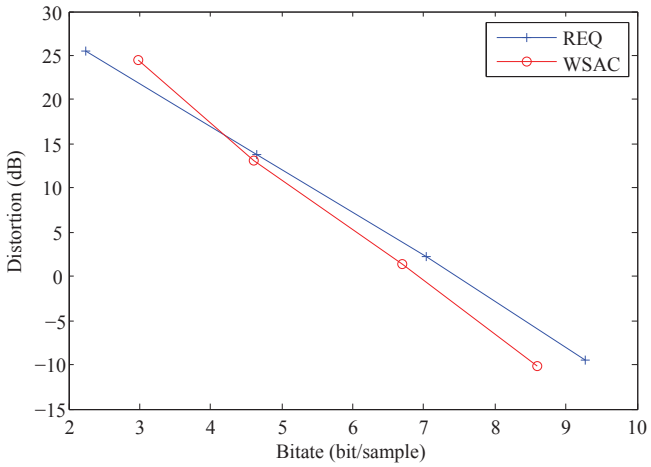

Fig. 7. Comparison of 4-layer WSAC and REQ ($L = 7$). 4-bit quantizers used for REQ and 5-bit quantizers for WSAC (except the 4-bit quantizer used for the last layer).

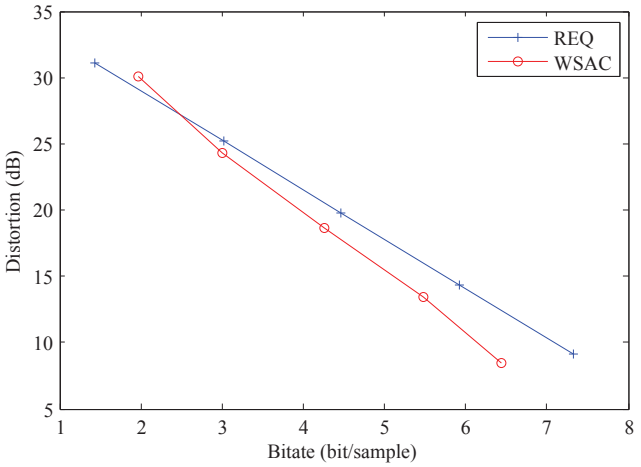

Fig. 8. Comparison of 5-layer WSAC and REQ ($L = 7$). 3-bit quantizers used for REQ and 4-bit quantizers for WSAC (except the 3-bit quantizer used for the last layer).

Figures 7 and 8 show the comparisons between WSAC and the REQ scalable coding systems in terms of bitrate-distortion performance. In the first one the 4-layer systems were compared and in the second one the 5-layer systems. The points show the possible rate-distortion pairs, that is for a 1-layer system , a 2-layer system and so on. As can be seen in these two figures when the number of layers increases the performance of WSAC becomes considerably better than REQ. The more number of layers used, the better WSAC performs compared to REQ. The only case where REQ works better is a

one layer system which is obvious since in that case a double-resolution quantizer is used in the base layer while there are no more layers and hence no savings for the other layers using QIM. In fact the proposed method starts to outperform when more than one layer is sent. Also comparing the two figures reveals that in the 5-layer systems, the amount of WSAC's outperforming increases since ΔB in (14) is smaller for a 3-bit quantizer compared to the 4-bit one.

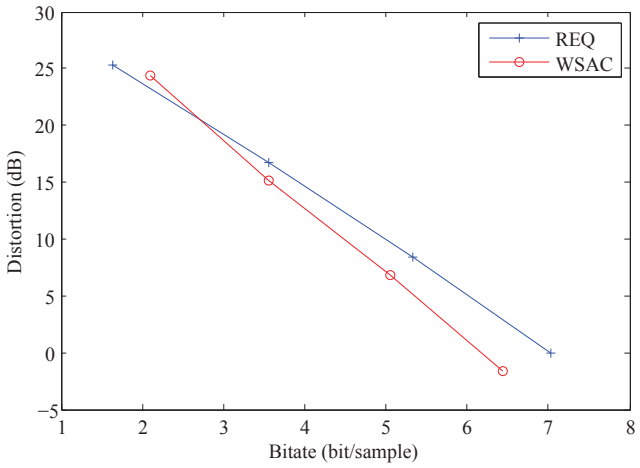

Fig. 9. Comparison of 4-layer WSAC and REQ ($L = 10$). 4-bit quantizers used for REQ and 5-bit quantizers for WSAC (except the 4-bit quantizer used for the last layer).

In Fig. 9 the performance of the 4-layer systems were compared for L=10. It can be seen that the performance difference between the two systems becomes higher for larger L or equivalently smaller input variance.

6. CONCLUSION

REQ scalable coding is a practical scalable coding scheme which is used in MPEG-4 audio coding. We proposed a modified version of such a system in which a watermarking technique known as QIM was used to embed some of the information of each layer in the previous one. The proposed method considerably outperforms REQ scalable coding in terms of rate-distortion and can be considered as a suitable replacement for practical scalable audio coders.

7. REFERENCES

[1] D. Ning and M. Deriche, "A bitstream scalable audio coder using a hybrid WLPC-wavelet representation," in *Proc. IEEE ICASSP*, Apr. 2003, vol. 5, pp. V–417–20.

[2] H. Huang, H. Shu, and S. Rahardja, "Bit-plane arithmetic coding for Laplacian source," in *Proc. IEEE ICASSP*, Mar. 2010, pp. 3358 –3361.

[3] D. H. Kim, J. H. Kim, and S. W. Kim, "Scalable lossless audio coding based on MPEG-4 BSAC," in *Proc. 113th AES Conv.*, Oct. 2002, Paper Number:5679.

[4] R. Yu, S. Rahardja, L. Xiao, and C. C. Ko, "A fine granular scalable to lossless audio coder," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1352 –1363, Jul. 2006.

[5] M. Movassagh, J. Thiemann, and P. Kabal, "Joint entropy-scalable coding of audio signals," in *Proc. IEEE ICASSP*, Mar. 2012, pp. 2961 –2964.

[6] R. Geiger, J. Herre, S. Kim, X. Lin, S. Rahardja, M. Schmidt, and R. Yu, "ISO/IEC MPEG-4 high-definition scalable advanced audio coding," in *Proc. 120th AES Conv.*, May 2006, Paper Number:6791.

[7] A. Aggarwal, S. L. Regunathan, and K. Rose, "Efficient bit-rate scalability for weighted squared error optimization in audio coding," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1313 –1327, Jul. 2006.

[8] B. Chen and G.W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. Inform. Theory*, vol. 47, no. 4, pp. 1423 –1443, May 2001.

[9] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 advanced audio coding," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, Oct. 1997.

[10] G. J. Sullivan, "Efficient scalar quantization of exponential and laplacian random variables," *IEEE Trans. Inform. Theory*, vol. 42, no. 5, pp. 1365 –1374, Sep. 1996.

[11] R. M. Gray and D. L. Neuhoff, "Quantization," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2325 –2383, Oct. 1998.