Title	Optimum quantizer design for pre-echo control in a low bit rate audio coding system(Accepted version)
Author(s)	Lim, Ai Chye; Huang, Dong Yan; Foo, Say Wei
Citation	Lim, A. C., Huang, D. Y., & Foo, S. W. (1999). Optimum quantizer design for pre-echo control in a low bit rate audio coding system. In Proceedings of the International Conference on Information, Communications and Signal Processing 1999: (pp.1-4). National University of Singapore.
Date	1999
URL	http://hdl.handle.net/10220/4670
Rights	© 1999 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder. http://www.ieee.org/portal/site

Optimum Quantizer Design for Pre-echo Control in a Low Bit-Rate Audio Coding System

Ai Chye Lim*, Dong-Yan Huangand Say Wei Foo***

*Department of Electrical Engineering
National University of Singapore
10 Kent Ridge Crescent
Singapore 119260
aichye@wildmail.com
elefoosw@nus.edu.sg

**Institute of Microelectronics
11, Science Park Road
Singapore Science Park II
Singapore 117685
dongyan@ime.org.sg

Abstract

In a low bit-rate audio transform coding system, the quantization noise is spread out over the entire analysis block in the time domain after the inverse transformation in the decoder. This increases the likelihood of pre-echoes occurring prior to a transient event. In order to control pre-echoes level, a suitable quantizer should be applied in the coder. This paper presents such an optimum quantizer design by a study of the input signal characteristic in terms of its dynamic range and probability density function (PDF) and a comparison study of different quantizers in a low constant bit rate audio coder with adaptive window switching technique. The performance of quantizers is measured by an objective mean i.e. the average segmental signal-to-noise ratio (SEGSNR).The Laplacian quantizer provides excellent results for the quantization of the transient signal with an increase of about 9 dB in the SEGSNR as compared to the non-uniform quantizers(MPEG-4).

1. INTRODUCTION

It is well known that the inappropriate temporal spread of quantization noise leads to a phenomenon called pre-echo in a low bit rate transform coding system, especially for transient signals. In this case, the quantization noise is not masked by the transient signal and becomes audible. Several techniques have been proposed to avoid pre-echo artifacts in the encoded/decoded signal [1][2][3][4][5].

Our intention is to take measures for handling pre-echo situation in a low bit rate audio coder based on Modified Discrete Cosine Transform (MDCT) technique. This coder employs a high frequency resolution filterbank together with a combination of pre-echo control, adaptive window switching, bit reservoir, gain control and the temporal noise shaping (TNS) technique. The first two techniques have been investigated and are outlined [6]. In order to take into consideration of the case of signal containing closely spaced transient events, a new intermediate transition window type and a method for an attack processing have been proposed for adaptive window switching technique. However, this technique cannot control quantization noise level prior to the attack of a sound. If the quantizer used is not adequate to the coder, even the noise is in the order of 0.5-2 ms prior to the transient event, the quantization noise may not be masked. This limitation of adaptive window switching leads to an optimum quantizer design and a comparison study of various quantizers such as uniform, non-uniform (MPEG4, A-law, μ-law) and Laplacian quantizers.

In order to design an optimum quantizer in the coder, the estimation of probability density function (PDF) of various music sources, especially the percussive sound (e.g., castanets and triangles) is carried out. The study suggests using a Laplacian quantizer in the coder. The simulation results report that average segmental signal-to-noise (SNR) by Laplacian quantizer is improved by about 9 dB over other quantizers for transient signals and at least about 0.9 dB for steady-state signal. In fact, this method increases the coding precision for the transform coefficients covering the transient signal portion. The informal subjective listening test shows that the pre-echo is controlled to a satisfactory level when the coder uses Laplacian quantizer with our proposed adaptive window switching approach [6].

In the following, firstly we present the basic structure of audio coding system studied in the paper. Then we discuss the optimum quantizer design problem. In section III, we present the simulation results and informal subjective test. Finally, we summarize our major work and outline our future work.

2. AUDIO CODING SYSTEM

The basic structure of audio codec used in this paper is presented in Figure 1. The input signal x(n) is transformed into frequency domain coefficients **X**(k) by MDCT based on time domain aliasing cancellation (TDAC) technique[7]. The coefficients **X**(k) are then quantized with a quantization step size selected according to the number of bits allocated to the coefficients. An adaptive bit allocation algorithm is used to derive the number of bits allocate to the coefficients [8]. In the decoder side, the quantized coefficients **X**'(k) are transformed back to the signal x'(k) in time domain by using the inverse MDCT. This technique leads the quantization noise from one transform block to spread out in time and becomes audible before a transient event, known as a "pre-echo phenomenon". In this case, we used an adaptive window switching approach proposed in [6] to reduce the pre-echo. This technique adapts the size of analysis block to the characteristics of the input signal. The objective measure (SNR) and subjective test show that the pre-echo is considerable reduced when we used a non-uniform quantizer (MPEG4). However, the quality of output is still distinguishable from original signal. As the quantization noise is the reason for the appearance of pre-echo using the transform coding technique in low bit rate audio codec. The controlling of quantization noise level seems an efficient method to reduce pre-echo. Based on this idea, an optimum quantizer design and a comparison study of different quantizers are carried out in this paper.

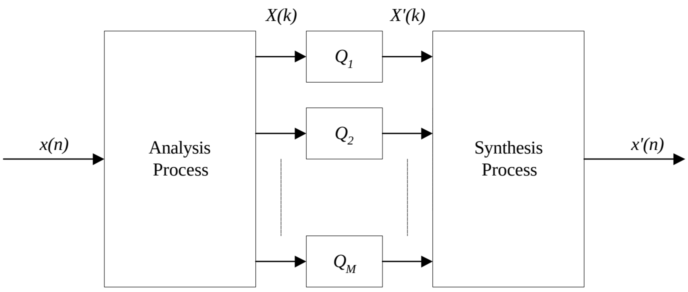

Figure 1: Framework for Analysis/Synthesis and Quantization Process

In the following , we discuss the quantizer design so as to provide an optimum quantizer for the transient signals given a fixed number of bits. The issue here is to determine a quantizer to optimize the bits allocated so as to improve the performance of the coder.

3. QUANTIZER DESIGN

Quantization is the conversion of a continuous-amplitude signal into a discrete-amplitude signal. The difference in amplitude between the quantized samples and the continuous-amplitude signal at the instant of sampling is known as the quantization noise [9]. When each of a set of the discrete values is quantized independently, the process is known as scalar quantization. Each of the analogue amplitude is compared against a finite set of amplitude values and the nearest value from the finite set is chosen to represent the analogue amplitude.

The quantization process is related to the bit rate. In general, quantization error variance is the most important quantity for comparing the performance of quantizers [10]. Let X be a zero mean random variable (i.e. m_x = 0) at the quantizer input with variance s_x^2, maximal absolute value x_{max} and probability density function (PDF) $p_x(.)$. The quantizer $Q(.)$ maps X into the discrete-values random variable Y. The quantization error Q = X – Y is also a random variable with PDF $p_q(.)$. The derivation of the A sou quantization error variance can be found in [10].

$$s_q^2 = \frac{1}{3} x_{\max}^2 2^{-2R} \qquad (1)$$

Equation (1) shows that the standard deviation of the noise decreases exponentially with increasing bit rate R. This explains why low bit rate leads to larger quantization error.

In consequence, the signal-to-noise ratio (SNR) can be derived as

$$SNR(dB) = 6.02R + c \qquad (2)$$

where c is a constant and depends on the quantizer. From above equation, we know that an increase of R by 1 bit implies an increase of about 6.02 + c in SNR.

As we know, low bit rate results in coarse quantization of samples. This is due to the fact that the number of bits available for the quantized levels is reduced such that the step size becomes larger. This leads to a larger quantization noise, which in turn affects the appearance of pre-echoes. A larger amount of noise will be spread in the time domain after the inverse transformation in the decoder. This increases the likelihood of pre-echoes occurring prior to a transient event. Hence, it is necessary to discuss the quantizer design so as to provide an optimum quantizer to match the input signal characteristic in terms of its dynamic range and probability density function (PDF) by a fixed bit budget, so that the quantization noise is as small as possible. Since there is only a fixed number of bits, the issue here is to design a quantizer such that it will optimize the bits allocated so as to give excellent performance of the coder, which can be measured by an objective measure i.e. the average segmental signal-to-noise ratio (SEGSNR).

In order to determine the optimum quantizer for the transient signal, it is necessary to estimate the PDF of the coefficients to identify any well-known distribution [11]. The study of the distribution of the MDCT coefficients can be done using the histogram, where each of the MDCT coefficients is treated as a random variable.

The histogram is a graphical display of the data such that the characteristic is subdivided into classes. It is a tabulation of the frequency of occurrence of the random variable in each classes. The relative frequency in each class is found by dividing the frequency in each cell by the total number of observations. The classes are of equal width and chosen to be adjacent. When there are outliners, i.e. values that are either too large or too small compared with the majority of the observations, the two end classes may be kept open-ended. To achieve a good estimation of the distribution of the coefficients, it involves some compromise in setting the class width, or rather the number of classes. If too few classes are chosen, specific details of the data are lost. On the other hand, if too many classes are selected, a summary of how the data is distributed will not be achieved. Large amount of data is required for accurate generation of the histogram.

The histogram of the MDCT coefficients for the music source castanets is plotted with the relative frequency (*y*-axis) in log-scale against the magnitude of the MDCT coefficients normalized with the standard deviation s_x of the coefficients (*x*-axis).

The curve is symmetrical and has its peak at the mean zero as shown in Figure 2. The gradient of the graph has an almost straight slope. Also, *P(x)* decreases with |*x*|. The number of classes is chosen to be 2 to the power of 10 to give good results. The result shows that most of the MDCT coefficients has small values clustered around the mean. More than 50% of the coefficients are distributed around the mean. Hence, both ends of the histogram are open-ended as there are outliners. The distribution of MDCT coefficients is found to best fit a Laplacian PDF.

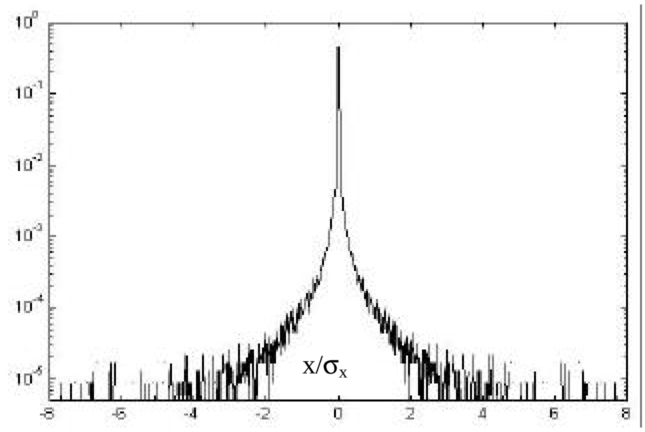

Figure 2: Histogram of MDCT Coefficients for Music Source Castanets

From above results, it is shown that the distribution of the MDCT coefficients for the transient signal (i.e. castanets) follows a Laplacian PDF. Hence, the Laplacian compressor characteristics should be considered [10]. The different quantizers used in various low-bit-rate audio codecs have been reported [12]. In order to design an optimum quantizer in audio codec in Figure 1, we implemented uniform, non-uniform (A-law, μ-law and MPEG4) and Laplacian quantizer. Our experimental results show that the MPEG4 non-uniform quantizer is better than the A-law and mu-law quantizers in audio coding systems. In the following simulation, we take into consideration the uniform, the non-uniform Equation (3) and the Laplacian quantizer Equation (4) to do a comparison study.

$$C(x) = x^{\frac{3}{4}} \qquad (3)$$

$$C_{opt}(x) = x_{\max}\,\mathrm{sgn}(x)\,\frac{1-e^{-\frac{h|x|}{x_{\max}}}}{1-e^{-h}}, \qquad h_{opt} = \frac{\sqrt{2}}{3s_x}x_{\max} \quad (4)$$

4. SIMULATION RESULTS

The uniform, the non-uniform and the Laplacian quantizers are implemented in the codec at 64 kbit/s.. The simulation has been carried for various monaural music .

Table 1 gives the segmental SNR of the coder for three quantizers, where the adaptive window switching technique is used. The segmental SNR for the three music sources using the uniform quantizer has the poorest quality both in the objective and subjective measure.

It is found that the Laplacian quantizer offers the best performance for the transient signal, with an increase of about 25.2% in segmental SNR over the non-uniform quantizer. This is consistent with the fact that the distribution of the MDCT coefficients of the castanets tends to follow a Laplacian PDF. The increase in segmental SNR of the music sources piano and organ is about 9.8% and 1.6% respectively, which is small compared to that of the castanets. But, there is still improvement. The informal subjective listening test also confirms that the pre-echo is reduced to satisfactory level.

Table 1: SEGSNR Comparison for Quantizers

SOURCE	Uniform	Non-uniform	Laplacian	Improvement in SEGSNR
	SEGSNR/dB			
Castanets	33.023	36.277	45.413	9.136
Piano	41.488	46.180	50.716	4.536
Organ	50.032	56.562	57.455	0.893

In practice, we can estimate the performance of Laplacian quantizer for different monaural music according Figure 3. From Equation (4), it can be deduced that the best η-law performance is achieved when the Laplacian input has the s_x value of $\sqrt{2}x_{\max}/3h$. Figure 3 shows the effect of a general η value on the SEGSNR for the music source castanets. This η value is not necessarily equal to the optimum. The η = 0 point refers to the case of uniform quantization. The curve for the SEGSNR did not follow a

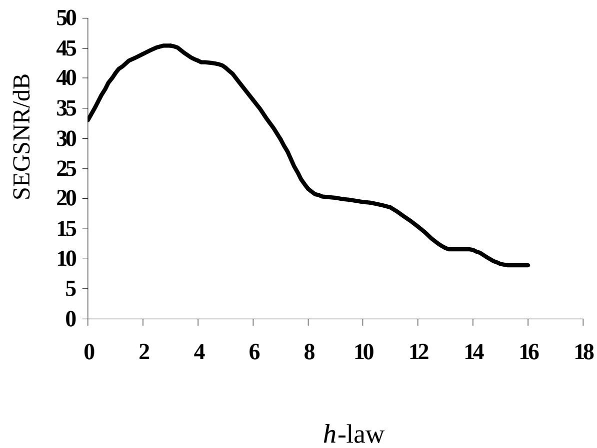

Figure 3: Performance of *h*-law Quantization

smooth increasing or decreasing curve. Instead, it increases from 33.023 dB at η = 0 to a maximum value of 45.413 dB at

η = 3, after which it decreases at different rates. This is due to the fact that the castanets contain transient events, such that the η same value cannot be used to model the variance of each frame.

5. CONCLUSIONS

The optimum quantization design in the coder (Figure 1) has been carried out. The simulation results and the informal subjective listening test show considerable improvement of audio quality in the pre-echo situation with Laplacian quantizer. This is because the optimum quantizer is to increase the coding precision for the transform coefficients covering the transient signal portion over other quantizers. In order to further enhance the coding precision, a bit reservoir technique [1] will be investigated.

REFERENCES

[1] Jürgen Herre and James D. Johnston, "Enhancing the Performance of Perceptual Audio Coders by Using Temporal Noise Shaping (TNS)", 101st AES convention, Los Angeles 1996, Preprint 4384.

[2] K. Brandenburg, "OCF – A New Coding Algorithm for High Quality Sound Signals", Proc. IEEE Conference on Acoustics, Speech and Signal Processing ICASSP–87, page 141–144.

[3] Hans Georg Musmann, "The ISO Audio Coding Standard", IEEE Global Telecommunications Conference and Exhibition, 'Communication: Connecting the Future', GLOBECOM '90, vol. 1, page 511–517, Dec 1990.

[4] Masahiro Iwadare, Akihiko Sugiyama, Fumie Hazu, Akihiro Hirano and Takao Nishitani, "A 128 kbits/s Hi-Fi Audio CODEC Based on Adaptive Transform Coding with Adaptive Block Size MDCT", IEEE Journal on Selected Areas in Communications, vol. 10, No. 1, Jan 1992.

[5] Akihiko Sugiyama, Fumie Hazu, Masahiro Iwadare and Taokao Nishitani, "Adaptive Transform Coding with an Adaptive Block Size (ATC–ABS)", IEEE Conference on Acoustics, Speech and Signal Processing, ICASSP–90, vol. 2, page 1093–1096, Apr 1990.

[6] Ai Chye Lim, "Handling Pre-echo Artifacts in a Low Bit-Rate Audio Coding System", Final Year Thesis, Department of Electrical Department, National University of Singapore 1999.

[7] John P. Princen and Alan Bernard Bradley, "Analysis/Synthesis Filter Bank Design Based on Time Domain Aliasing Cancellation", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. ASSP–34, No. 5, page 1153–1161, Oct 1986.

[8] Allen Gersho and Robert M. Gray, "Vector Quantization and Signal Compression", Kluwer Academic Publisher, page 233–235, 1992.

[9] A. M. Kondoz, "Digital Speech: Coding for Low Bit Rate Communication System", John Wiley & Sons, page 7-13, 1994.

[10] N.S. Jayant and Peter Noll, "Digital Coding of Waveforms – Principles and Applications to Speech and Video", Prentice Hall Inc, Englewood Cliffs, New Jersey, page 31–35, 1984.

[11] Kok Keong Chan, "Adaptive Transform Coding of Speech Signals At 16 kbits/s and Below", Final Year Thesis, Department of Electrical Engineering, National University of Singapore, 1983.

[12] Karlheinz Brandenburg and Marina Bosi, "Overview of MPEG Audio: Current and Future Standars for Low-Bit-Rate Audio Coding" tutorial on MPEG/Audio Compression," Journal Audio Eng. Soc., Vol. 45, No. 1/2, 1997 January/February.