DIFFERENT QUANTISATION NOISE SHAPING METHODS FOR PREDICTIVE AUDIO CODING

Stefan Wabnik, Gerald Schuller, Jens Hirschfeld, Ulrich Kraemer

Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany
{*wbk,shl,hfd,kre*}@*idmt.fraunhofer.de*

ABSTRACT

The paper presents a comparison of a previous and a new approach to shape quantization noise in low bit rate predictive audio coding. The previous approach uses an adaptation of the step size of a uniform quantizer, the new approach uses a quantizer with clipping. Both approaches are evaluated using a predictive audio coding scheme. The presented results of a listening test show the improved performance of the new approach.

1. INTRODUCTION

Our goal is a low bit rate audio coder with low delay, with as little annoying distortion as possible. To obtain the low delay goal we use a predictive audio coder, the so-called Ultra Low Delay coder (ULD) as described e.g. in [1]. It uses psycho-acoustically controlled linear filters for the quantization noise shaping.

Because of its structure, the quantization noise is always at the given threshold, even if there is no signal in a given frequency range. This noise remains inaudible as long as it corresponds to the real masking threshold. To obtain a bit rate which is even lower than the bit rate imposed by this threshold, it needs to be elevated, and hence the noise becomes audible. It becomes especially noticeable in areas where there is no signal part. Examples are the very low and the very high audio frequencies. In these ranges, there are usually very little signal parts, but a high masking threshold. If the masking threshold is uniformly elevated over the entire frequency range, the quantization noise is at the now elevated threshold, even if there is no signal, and becomes audible as an annoying artificial sounding signal. Subband based coders don't have this problem because they simply quantize a subband with smaller signals than the threshold to zero.

To avoid this problem of audible quantization noise where there is no signal, we propose a modified way to increase the quantization levels above the masking threshold. Instead of uniformly lifting the masking threshold, we add a certain level of the signal spectrum to it. In this way, there is no audible quantization noise where there is no signal.

This paper compares the old approach of uniformly elevating the masking threshold, and the approach of adding a certain level of the signal spectrum. Both strategies have been implemented in the Ultra Low Delay (ULD) audio coding scheme. The paper presents results of a listening test which was conducted to evaluate both strategies.

The rest of the paper is organized as follows: Section 2 gives a description of the audio coding scheme used, Section 3 describes the suggested noise shaping methods, Section 4 presents results of the conducted listening test, Section 5 gives some conclusions.

2. THE USED AUDIO CODER

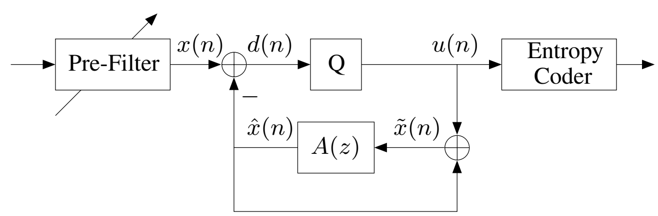

Fig. 1. Design of the coder with rate loop.

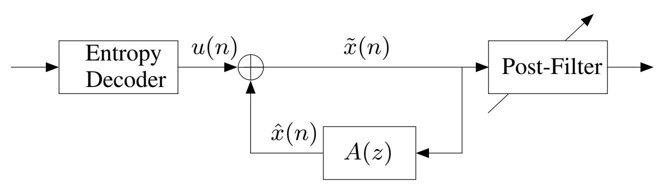

Fig. 2. Design of the ULD Decoder.

For some applications, not only high compression ratios, but also very low encoding and decoding delay has become an essential prerequisite. In live productions or in distributed productions where artists perform simultaneously in different studios the tolerable total coding delay is less than ten milliseconds.

The Ultra Low Delay Audio Coder achieves a total encoding/decoding delay of 5.33 to 8 milliseconds with sampling frequencies from 32 kHz to 48 kHz by separating the two aims of irrelevance and redundancy reduction and assigning

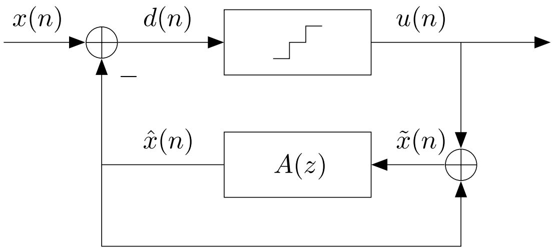

Fig. 3. Predictive encoder structure, Method I and II.

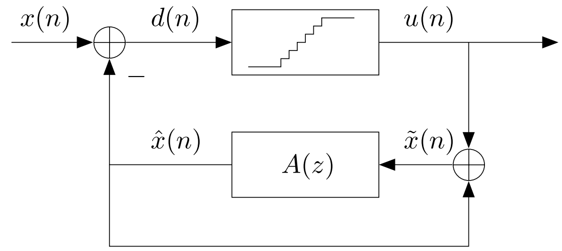

Fig. 4. Modified predictive encoder structure, Method III.

them to different functional units [2][3]. At the unchanged masking threshold it achieves bit rates in the range of 80 to 96 kbit/s. Fig.1 shows the main functional blocks of the encoder.

A psycho-acoustically controlled adaptive linear filter is used on the input audio signal for the irrelevance reduction. The perceptual model uses a DFT of length 256 with 50% overlap, which results in a delay of 128 samples. The estimate of the masking threshold is used to calculate the model parameter of an AR-model. The pre-filter uses this parameters to normalize the input signal with respect to the masking threshold. Compared to input signal, the pre-filtered signal is much smaller in magnitude.

Redundancy reduction is achieved via adaptive prediction and entropy coding. To tune the scheme to a certain bit rate, a gain factor is used which controls the magnitude of the pre-filtered signal, which is equivalent to changing the step size of the quantizer. The scale factor is adapted once every 128 samples, quantized and transmitted to the decoder as additional information.

From a pre-filtered signal $x(n)$, the signal $\hat{x}(n)$ is subtracted, and the difference $d(n)$ is quantized using a uniform quantizer $Q\{\cdot\}$ [4]. The signal $\tilde{d}(n)$ is transmitted to the decoder. The predicted signal $\hat{x}(n)$ is generated from a linear prediction filter $A(z) = \sum_{i=1}^{N} a_i z^{-i}$ of length N and depends only on quantized values. The coefficients a_i are calculated adaptively, either on a block-by-block basis of $x(n)$, or on past values of $\tilde{x}(n)$. The decoded signal is equal to signal $\tilde{x}(n)$ in the encoder. If the quantization is modeled as additive uncorrelated white noise $\epsilon(n)$, the following condition holds: $\tilde{x}(n) = x(n) + \epsilon(n)$.

The decoder (Fig.2) contains an entropy decoder, followed by the predictive decoder structure and a post-filter. The post-filter transfer function is the inverse of the pre-filter transfer function, and hence has a frequency response like the masking threshold. The added quantization noise is filtered by the post-filter, too, so the post-filter colors the added noise like the masking threshold.

3. DESCRIPTION OF INVESTIGATED NOISE SHAPING METHODS

In the following, the three investigated noise shaping methods are described.

3.1. Method I

For higher bit rates, the ULD coding scheme uses backward adaptive prediction, that is, the coefficients for the prediction filter $A(z)$ are updated from previously decoded signal values on a sample-by-sample basis. A quantizer with variable step size is used (see Fig.3). The step size is adapted every 128 samples using information from the entropy coder, and is transmitted as side information to the decoder. Method I increases the quantization step size, hence adds more white noise to the pre-filtered signal and so uniformly elevates the masking threshold.

3.2. Method II

Method II uses forward block-adaptive prediction, that is, the coefficients for the prediction filter $A(z)$ are calculated once every 128 samples from the unquantized pre-filtered samples and transmitted as side info (see Fig.3). The quantizer step size is adapted every 128 samples using information from the entropy coder, and is transmitted as side information to the decoder. This method we call ”method II”. Method II increases the quantization step size as method I. The difference to method I is, that the predictor update is not influenced by any quantization.

3.3. Method III

Method III uses forward block-adaptive prediction, as method II. But here, the quantizer has a given number N of steps with fixed step size (see Fig.4). For pre-filtered signals $x(n)$ with amplitudes outside the quantizer range $[-N\Delta; N\Delta]$ the quantized signal is limited to $[-N\Delta; N\Delta]$. This results in a quantization noise with a power spectral density psd_ϵ which is no longer white, but resembles psd_x, the power spectral density of its input signal, the pre-filtered audio signal. As

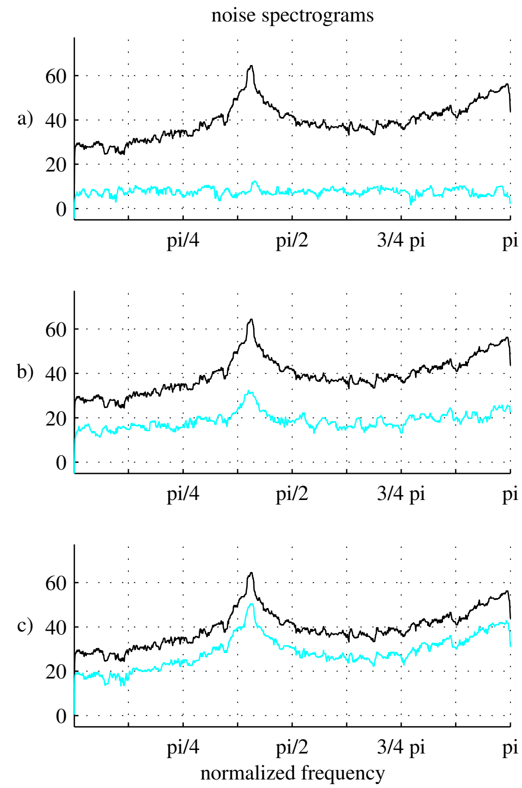

Fig. 5. Signal psd (upper graph) and error psd (lower graph) for different clipping ranges in method II: a)$[-15; 15]$, b)$[-7; 7]$, c)$[-1; 1]$.

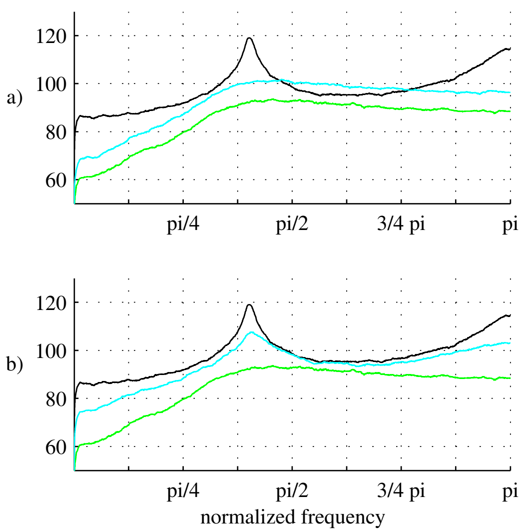

Fig. 6. Signal psd, error psd and masking threshold for a) method II, b) method III.

an example, each plot in Fig.5 shows the psd of an input signal $x(n)$ and the psds of the quantization error for different numbers of quantization steps (plots a) to c)). The signal $x(n)$ is colored noise with a power of $\sigma^2 = 34$. When quantized with step size $\Delta = 1$, the signal lies within $[-21; 21]$. For the plots a) to c), the quantizer range has been limited to $[-15; 15]$ in a), $[-7; 7]$ in b) and $[-1; 1]$ in c). The quantization error has been measured as the difference between the unquantized pre-filtered signal and the decoded pre-filtered signal. For clarity, the psd graphs of the error are plotted with an offset of -10dB. Method III adds quantization noise to the pre-filtered signal which resembles the power spectral density of the pre-filtered signal, depending on the severity of the applied clipping. Thus, after applying the post-filter in the decoder, the resulting noise psd is not elevated above the masking threshold where there is no signal. This is shown in Fig.6, plot b).

4. LISTENING TEST

All three noise shaping methods were tested with the ULD coding scheme in a listening test according to the MUSHRA standard [5] where anchors were omitted. The MUSHRA test was implemented on a Laptop computer with external DA-converter and STAX amplifier/headphones in a quiet office environment. The group of eight test listeners consisted of expert and non-expert listeners. Before the subjects started with the listening test, they had the possibility to listen to a test set.

The tests were conducted with 12 mono audio files of the MPEG test set: es01 (Suzanne Vega), es02 (male speech, German), es03 (female speech, English), sc01 (trumpet), sc02 (orchestra), sc03 (pop music), si01 (cembalo), si02 (castanets), si03 (pitch pipe), sm01 (bagpipe), sm02 (glockenspiel), sm03 (plucked strings). The audio files, with a sampling frequency of 32 kHz, were coded at a bit rate of 64 kbit/s. For method I, the gain range was $[0.01; 10.0]$ and a backward adaptive golomb coder was used for entropy coding, for method II, the quantizer step size was $\Delta = 1.0$ and the quantizer range was $[-1; 1]$.

The results of the MUSHRA listening test are presented in Fig.7, including 95%-confidence intervals as bars. As long as the confidence intervals overlap, there is no statistical significant difference between the coding methods. For item si03 (pitch pipe), method II is rated significantly lower than method III. For sm02 (Glockenspiel), method III is rated significantly better than the other two. The overall score (all items) also shows that noise shaping method III is rated significantly better than the other two methods. Furthermore, only method III was rated "good audio quality" under the given test conditions.

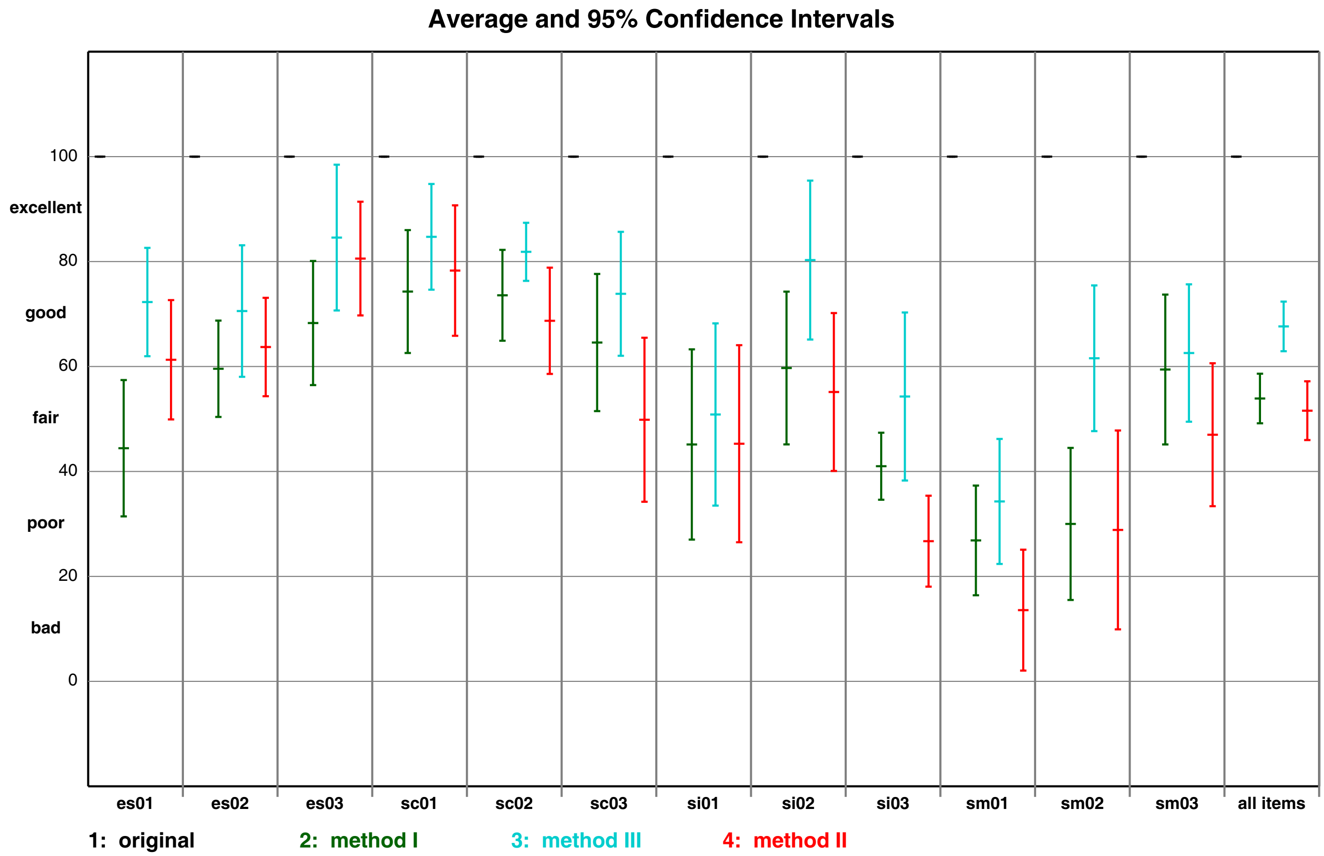

Fig. 7. Listening test results for the different setups in the listed order

5. CONCLUSIONS

We obtained our goal to find an improved quantization scheme for a low bit rate predictive audio coder with using a clipping uniform quantizer. With the listening test we found that it leads to fewer annoying artifacts than a quantizer with increased step sizes produces. An explanation is, that the clipping quantizer increases the quantization noise only at frequencies where there are also signal parts. This contrasts the behavior of a quantizer with increased step sizes, which increases its quantization errors uniformly across frequency, and hence leads to a uniformly elevated masking threshold in the predictive audio coder, also at frequencies where there are no signal parts.

6. REFERENCES

[1] M. Lutzky, G. Schuller, M.Gayer, U. Kraemer, and S. Wabnik, "A guideline to audio codec delay," *116th AES convention*, May 2004, Berlin, Germany.

[2] B. Edler, C. Faller, and G. Schuller, "Perceptual Audio Coding Using a Time-Varying Linear Pre- and Post-Filter," *109th AES convention*, September 2000, Los Angeles, CA, USA.

[3] G. Schuller and A. Härmä, "Low Delay Audio Compression using Predictive Compression," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2002, Orlando, FL, USA.

[4] N.S. Jayant and Peter Noll, "Digital Coding of Waveforms," *Prentice-Hall Signal Processing Series*.

[5] ITU-R BS.1534-1, "Method for the subjective assessment of intermediate quality levels of coding system," January 2003.