CODE EXCITED SAMPLE-BY-SAMPLE GAIN ADAPTIVE CODING FOR LOSSLESS COMPRESSION OF AUDIO SIGNALS

Yongmin LI and Cheung-Fat CHAN

Department of Electronic Engineering, City University of Hong Kong
Tat Chee Avenue, Kowloon, Hong Kong SAR
phone: + (852) 27887193, email: yongminli2@student.cityu.edu.hk, itcfchan@cityu.edu.hk

ABSTRACT

A coding algorithm for lossless compression of audio signals is presented. The proposed algorithm consists of a lossy coding part and a lossless coding part. The lossy coding part is based on code excitation approach where the excitation gain and the short-term prediction coefficients are adapted in a sample-by-sample fashion to cope with rapid time-varying nature of audio signals. The error between the input and the code-excited synthetic signal is then encoded by an arithmetic coder to achieve lossless compression. The excitation codebook is searched by using an M-L tree search strategy with minimum error energy and minimum code length after arithmetic coding as search criteria. The proposed coder has very low decoding complexity due to its simple code excitation structure and achieves compression performance comparable to other advanced lossless coders for coding CD quality audio.

1. INTRODUCTION

Audio coding has been an active research area for many decades. Many popular audio coders such as MP3 (MPEG-I Layer 3) and AAC are lossy coders which are based on exploring the psychoacoustics property of human auditory perception to achieve high compression ratios [1][2]. These coders, however, may introduce coding artifacts to the decoded audio signals even though most of these artifacts are perceptually inaudible. Recently, most efforts in audio coding research are directed to lossless compression of high quality audio where the audio data can be faithfully reproduced at the decoder without any distortion. There are numerous applications that require storage, transmission and processing of high-fidelity audio without distortion, for examples; music archival systems, distribution of studio audio materials, further processing of professional audio, and music broadcasting over the internet. High-end consumer applications such as home theater systems also demand high-fidelity audio reproduction. With the rapid proliferation of large capacity storage devices and high-speed internet connections, more applications utilizing lossless audio compression technology are expected to appear. Recent research trends in lossless audio compression focus on a decorrelation-entropy coding approach in which the audio signal is firstly gone through a decorrelation process. The decorrelated signal has a smaller entropy which can then be efficiently encoded by a lossless entropy coder such as arithmetic coder [3][4], for example; an IIR filter is used in MLP coding for DVD-Audio to decorrelate the audio signal prior to arithmetic coding [5]. The main disadvantage of this approach is that the decorrelator has to be implemented using lossless arithmetic which requires high precision fixed-point processing. In addition, the order of this decorrelation filter is generally very high and to compute the filter coefficients on the fly in real time is quite a demanding task both in the encoder and decoder. Most implementations only employ finite sets of filter coefficients pre-computed and stored in a table. During encoding, the best set is selected from the table to avoid real-time computation from the input data and also to guarantee filter stability [5]. In this paper, a new lossless coding algorithm is proposed. Instead of using a decorrelator, the proposed algorithm employs a code-excited linear predictive strategy to code the audio waveform as close to the original as possible, and then the residual having smaller entropy is further encoded by an arithmetic coder.

2. METHODOLOGY

A block diagram of the proposed coder is shown in Fig. 1. The proposed lossless coder is built upon a highly efficient lossy coding part and a lossless entropy coding part, where the error between the input signal and the synthetic signal from the lossy coding part is encoded by an entropy coder. The coding parameters for the lossy part and the entropy-encoded residual are sent to the decoder to achieve lossless compression. In the proposed coding system, the lossy coding part is based on code excitation strategy similar to code-excited linear predictive (CELP) coder for coding speech signals [6]. However, they are fundamentally different in many aspects. The CELP algorithm for speech coding uses an adaptive codebook for modeling the periodicity of voiced speech. This adaptive codebook is constructed from the past excitation signal with a shift delay equal to the pitch of voice speech. But this single-tap pitch adaptive codebook may not be effective for coding general audio signals such as music which may have multiple tones generated from various musical instructions. Moreover, the abstract form of the parameters for the adaptive codebook, i.e., a pitch lag and a pitch gain, is insufficient to achieve accurate matching to the input signal such that the residual signal would have a

 ISSN 2076-1465

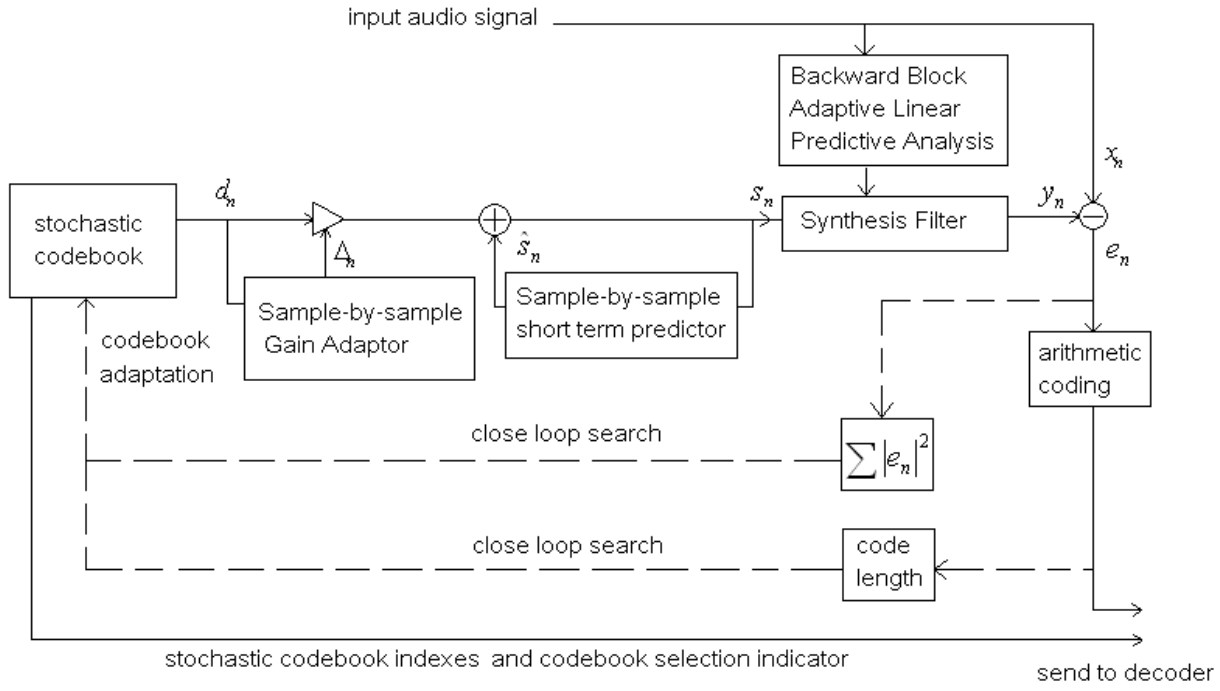

Fig.1. Block Diagram of the Proposed Lossless Audio Coder

large variance which is not beneficial to entropy coding in the later part. Therefore, the proposed excitation structure does not utilize an adaptive codebook, instead, only a simple stochastic codebook is used and the excitation gain of the stochastic codeword is adapted in a sample-by-sample fashion. Note that in conventional CELP coding of speech, the stochastic vector is multiplied by a gain which is unchanged within the subframe period. In the proposed system, after scaling by the adaptive gain factor, the scaled excitation signal is then added with the output from a short-term linear predictor whose coefficients are also adapted in a sample-by-sample fashion. This sample-by-sample adaptation approach can model rapid amplitude variation of the source waveform and remove inter-sample correlation effectively. Due to the complexity of close-loop searching the excitation codebook, the order of this sample-by-sample adaptive predictor must be low, therefore, in order to capture the steady correlation at a longer term, a high order linear predictive synthesis filter which is backward-adapted from the input signal in a frame-by-frame basis is further applied to generate the synthetic signal. Finally, the error between the input and the synthetic signal is encoded by an arithmetic coder to complete the lossless coding system. The arithmetic-encoded error signal, the optimum code indices of the stochastic codebook and the codebook selection indicator are the information sent to the decoder. Detailed explanations for each coding block of the proposed coder are given as follows:

2.1 Excitation Codebook

The excitation codebook is a fixed codebook constructed from a collection of stochastic codewords. Each stochastic codeword is a vector of D elements. Each element is a signal sample d_n having amplitude limited in the range of ± 1. If the codebook has B stochastic vectors, the bit rate required to encode an audio sample for the lossy coding part of the system is $(\log_2 B)/D$ bps. The stochastic codewords are obtained by training from a large collection of real audio signals. The training process is an iterative procedure. An initial codebook is constructed from uniformly distributed random codewords. The training vectors are partitioned into quantization clusters corresponding to the excitation codewords by going through the encoding process. Then, each excitation codeword is re-optimized from its cluster of training vectors according to the principle of minimum coding error energy. This is done by minimizing the error energy ε_m within the same cluster C_m with respect to d_n,

$$\varepsilon_m = \sum_{x_n, y_n \in C_m} \sum_{n=1}^{D} (x_n - y_n)^2 \tag{1}$$

where x_n is the input signal and y_n is the synthetic signal obtained as

$$y_n = (\hat{s}_n + d_n \Delta_n) * h(n) \tag{2}$$

where d_n is the excitation signal, Δ_n is the excitation gain, $\hat{s}_n$ is the predicted sample of s_n, i.e., the input signal to the block-adaptive synthesis filter, and $h(n)$ is the impulse response of the block-adaptive synthesis filter. Assuming each element of the excitation vector can be optimized independently and by setting

$$\frac{\partial}{\partial d_n}\left[\sum_{x_n, \hat{s}_n, \Delta_n, h(n) \in C_m} \left(x_n - (\hat{s}_n + d_n \Delta_n) * h(n)\right)^2\right] = 0, \; n = 1,2..D \tag{3}$$

then d_n can be re-estimated during codebook training as

$$\bar{d}_n = \frac{\sum_{x_n,\hat{s}_n,\Delta_n,h(n)\in C_m} \left(x_n - \hat{s}_n * h(n)\right)\left(\Delta_n * h(n)\right)}{\sum_{\Delta_n,h(n)\in C_m} \left(\Delta_n * h(n)\right)^2}, \quad n = 1,2..D \tag{4}$$

It is found that about 1.5 dB improvement in SNR can be achieved with a trained codebook.

2.2 Sample-by-sample Gain Adaptation

Conventional CELP coding algorithm employs a fixed excitation gain for the whole subframe period. The gain has to be quantized and sent to the decoder. For speech coding using large subframe size the overhead is small, but for audio coding, this approach is not feasible because the frame size must be small, typically less than 5 samples, in order to get accurate matching of fast time-varying audio signals. Moreover, the overhead in sending a fixed gain for every small frame is too large. In waveform coder such as ADPCM [7], the difference between the input and the predicted signal is scaled by a scaling factor and the scaled difference signal is then applied to a quantizer for quantization. The scale factor is adapted from the quantized codeword so that if the difference signal is large, the scale factor is increased to reduce the magnitude of scaled difference signal and hence pulls the signal back to the quantization range [8]. In this paper, a new adaptation mechanism based on a concept similar to scale factor adaptation of ADPCM coder is proposed to adapt the excitation gain in a sample-by-sample fashion directly from the excitation codeword. In this method, the excitation gain Δ_n is adapted as:

$$\Delta_{n+1} = \left(\beta + |d_n|\right)^{\alpha} \cdot \Delta_n \tag{5}$$

where d_n is the excitation signal having magnitude limited to 1 β is a threshold value chosen such that if $|d_n| > 1-\beta$, the gain is increased, otherwise it will be decreased. The rate of increasing or decreasing is an exponential function of time and α is a modification factor to further control the adaptation rate. This adaptation formula is very simple but still capable of following rapid increases in signal magnitude during musical attacks and also allows smooth decaying during musical releases. The main reason for using such a simple adaptation formula is that this gain adaptation has to be performed for each codeword from the excitation table during close-loop codebook search, and the search complexity will be extremely high if a complicated rule is used. Intuitively, the threshold value β is considered to be at the midpoint of the signal magnitude range, i.e. 0.5, however, after an optimization procedure similar to codebook training described in previous session is applied, an optimum value of $\beta = 0.727$ is obtained. That means if $|d_n| > 0.273$, the next gain value should be increased. The optimization is based on minimizing the error energy from a large collection of audio pieces including pop music and classical music. All of these audio signals are sampled at 44.1 kHz with 16 bit sampling rate and the amplitude range is ± 1. This finding is a bit surprising, and possible explanation is that audio signals generally have non-uniform distribution; the statistical mean of their magnitudes is probably at around 0.273 instead of 0.5. Furthermore, various kinds of music have been run to determine the modification factor of the adaptation rate, and $\alpha = 1$ is found to be a good compromise for its lowest adaptation complexity. With this sample-by-sample gain adaptation strategy, the use of a long-term adaptive codebook is not necessary.

2.3 Sample-by-sample Adaptive Short Term Predictor

In order to remove inter-sample correlation to further reduce the error signal, a short-term predictor is necessary. The predictor should have sufficiently high order to capture short-time correlation while allowing rapid adaptation to cater for time-varying characteristics of audio signals. Of course, the best performance can be achieved by a high-order predictor with its coefficients adapted in a sample-by-sample manner. However, the encoding complexity will be extremely high because the coefficient adaptation process has to be done within the codebook search loop. In this work, a hybrid approach using two predictors is applied. First, an order-2 predictor is applied in the coding path and its coefficients are adapted in a sample-by-sample manner. Second, a high-order linear predictive synthesis filter with its coefficients adapted on a frame-by-frame basis is applied later. The coefficients of the order-2 predictor are adapted within the codebook search loop, while the high-order predictor does not. The output from the order-2 predictor is

$$\hat{s}_n = a_1(n)s_{n-1} + a_2(n)s_{n-2} \tag{6}$$

where $a_1(n)$ and $a_2(n)$ are the predictor coefficients. By minimizing the prediction error energy

$$\varepsilon = \sum_n \left[s_n - \hat{s}_n\right]^2 \tag{7}$$

with respect to the predictor coefficients, the optimum prediction coefficients can be computed as:

$$\begin{bmatrix} a_1(n) \\ a_2(n) \end{bmatrix} = \frac{1}{r_0(n-1)r_0(n-2) - r_1^2(n-1)} \begin{bmatrix} r_0(n-2)r_1(n) - r_1(n-1)r_2(n) \\ r_0(n-1)r_2(n) - r_1(n-1)r_1(n) \end{bmatrix} \tag{8}$$

where the autocorrelations are computed recursively as

$$r_0(n) = 0.98 \times r_0(n-1) + s_n s_n \tag{9}$$

$$r_1(n) = 0.98 \times r_1(n-1) + s_n s_{n-1} \tag{10}$$

$$r_2(n) = 0.98 \times r_2(n-1) + s_n s_{n-2} \tag{11}$$

It is found that the prediction performance achieved by directly computing the prediction coefficients using equation (8) is much better than method used in ITU G.726 ADPCM coder which recursively updates the filter coefficients using LMS algorithm. The determined $a_1(n)$ and $a_2(n)$ are always stable, because a close-loop searching method is used in the encoder to guarantee the minimum error energy. Therefore if $a_1(n)$ and $a_2(n)$ are unstable, the corresponding codeword will not be selected.

2.4 Backward Block Adaptive Linear Predictive Filter

To remove more short-term correlation over a longer period, a high-order linear predictor is necessary. By accompanying the aforementioned sample-by-sample short term predictor to follow the changing of input signals in more details, the order of this block adaptive linear predictor does not need to be very high. Here, an order-10 lattice synthesis filter is used. The reflection coefficients are computed from the past

input samples in a frame-by-frame basis by using an asymmetric window. The frame shift is 64 samples. These filter coefficients are fixed during the codebook searching process so the complexity involved is relatively small as compare to the sample-by-sample adaptation. Since the filter is backward adaptive, its coefficients do not need to be sent to the decoder, because the coding scheme is lossless.

2.5 Excitation Codebook Search criterion

The search process comprises two stages: codewords search and adaptive codebook selection.

2.5.1 codewords search

During encoding, a full search technique is performed to select the "best" excitation codeword from the stochastic codebook. In order to reduce the search complexity, the codebook size and the codevector dimension must be kept small. A number of codebooks with different combinations of dimensions and sizes have been designed, for examples; a 2 bit per sample (bps) codebook can have a codevector dimension of 2 and a codebook size of 16, (i.e., 4 bits for coding 2 samples), a 2.5 bps codebook can have a dimension of 2 and a size of 32. Since the codevector dimension is small, the inter-frame correlation can not be explored if the search is independent of its adjacent frames. In this work, the encoding frame size is 64 samples, and each frame is equally divided into 4 subframes. In each subframe, an M-L tree search technique is applied across several steps to improve the coding performance. The idea is to keep the best M codewords in each search step, and with the search depth of L steps, the best two paths are kept for extended M-L tree search in the next subframe while other paths are released after L-step search depth is reached. As a compromise of search complexity and performance, M=2 and L=4 are used. Then the M-L tree search is extended into the second subframe; four best two paths are gotten after the second L-step search depth is reached. So and so, at the end of one frame, 16 best paths are figure out and one of them is chosen as the best path to encode the whole frame. Because the codebooks have two kinds of dimension; when the dimension is four, the searching process is just as shown in Fig. 2. Otherwise, the dimension equals 2, which makes the process repeat once in each frame. But what is the criterion for choosing the best path from the 16 paths to reach our demand? Obviously, since the proposed coding system is a cascade of a lossy coding part and a lossless coding part, two codebook search criteria are possible. For the lossy coding part, minimum error energy can be used as search criterion. For lossless compression, the ultimate performance measure is only the encoding rate because there is no audio quality issue; therefore, code length after entropy coding can be used as a search criterion. However, since the code length is dependent on the entropy of the source and the entropy is related to the statistical distribution which is a very long term measure, entropy alone can not be used as codebook search criterion because the codebook search is done locally with short interval, therefore, we propose a combined minimum error energy and minimum entropy search criterion. For M-L tree search in each subframe, the search criterion is minimum error energy. After dealing with a whole frame, the residual signal as a result of each coding path is arithmetic-encoded, the codewords in a path that result in the smallest code length is fetched as the best codewords. Since M-L tree search using minimum error energy measure already guarantees small residual signal with small entropy, the combination of energy/entropy search can achieve better performance.

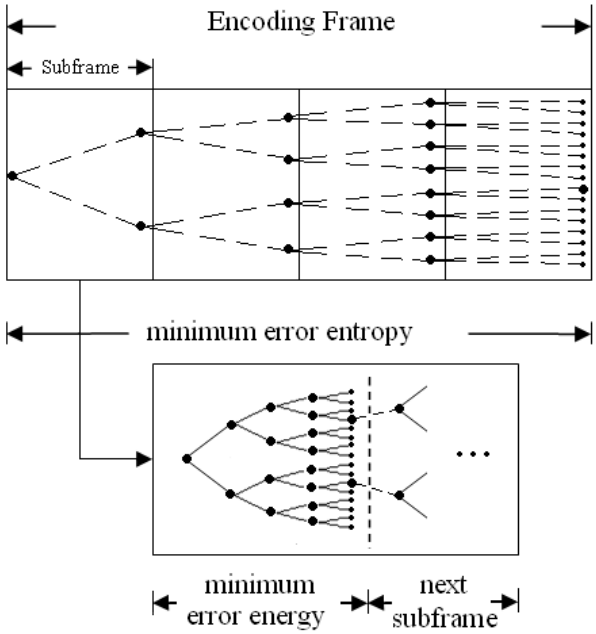

Fig. 2. Structure of Excitation Codebook Search

2.5.2 adaptive codebook selection

Since the encoding bit rate is the sum of the average bit rate after arithmetic coding of the residual and the bit rate for coding the codebook indices, which in terms is a function of the residual entropy, an adaptive choice of codebooks is more beneficial so as to account for various signal statistics. A total of nine possible combinations of codebooks have been tried in turn for every frame separately which have encoding rates ranging from 0.75 bps to 4.5 bps. And the adopted codebook is the one that has the smallest encoding bit rate. In other words, the codebook is adapted in every frame based on minimizing total entropy. From the observed statistic, four of the nine codebooks are more frequently selected. To save the overhead information telling the decoder which codebook is chosen, in the proposed codec only four codebooks are the candidate codebooks.

2.6 Arithmetic Coding

After code-excitation synthesis, the error between the input and the synthetic signal has low energy and hence small entropy. An adaptive arithmetic coder is then applied to code the error signal losslessly. The symbol probabilities of the error signal are computed using a sliding window approach and updated for each input symbol. The probability table represents the statistics of the past 500 samples; this provides the best performance for accurately matching to the local statistics of the input signal. The arithmetic coder is

implemented using an incremental shift out algorithm with low complexity.

3. EVALUATION RESULTS

The performance of proposed coding system is evaluated by encoding various audio pieces composed of pop music, jazz music, and classic music. All audio signals are sampled at 44.1 kHz with 16 bit A/D conversion. For lossy coding part, two performance indexes are used for evaluation, i.e., SNR and entropy of the residual signal. Table 1 lists the SNRs and entropies obtained from encoding three music pieces by the proposed codec.

For lossless coding part, two popular lossless audio coders; FLAC [9] and MPEG-4 ALS (RLS-LMS) [10], are used for comparison. Table 2 shows the average compression ratios achieved by these coders and our proposed coder for various kinds of audio pieces.

Table 1. List of the SNRs and entropies achieved by the proposed codec.

	Jazz	Pop	Classic
Entropy(bit)	6.013	7.648	8.080
SNR(dB)	48.445	28.835	31.466

Table 2. Comparison of compression ratios achieved by FLAC -2, MPEG-4 ALS (RLS-LMS) and our proposed coder.

	FLAC -2	MPEG-4 ALS	Proposed codec
Jazz	1.675	1.833	1.676
Pop	1.399	1.543	1.419
Classic	1.382	1.469	1.372
average	1.485	1.615	1.489

Under evaluation with compression ratio, the proposed coder performs slightly better than FLAC -2 and it only performs 7.80% worse than MPEG-4 ALS (RLS-LMS). Nonetheless the proposed coder has advantage of low decoding complexity as well as its lossy coding functionalities.

The decoding time of the proposed decoder is listed in Table 3 to compare with MPEG-4 ALS (RLS-LMS) decoder.

Table 3. Comparison of decoding time (in seconds)

	MPEG-4 ALS	Proposed codec	time saving
Jazz	16.94	8.27	51.18%
Pop	8.55	4.74	44.56%
Classic	16.45	8.95	45.59%

The decoding time of the proposed decoder is 47.64% less than that of MPEG-4 ALS (RLS-LMS) decoder, while the proposed decoder has not been optimized for speed.

4. CONCLUSION

A new lossless audio coding algorithm is proposed. The algorithm is a cascade of a lossy coding part and a lossless coding part. The lossy coding part is based on a modified code-excitation approach where the excitation gain and short-term predictor coefficients are adapted in a sample-by-sample fashion. A close-loop codebook search strategy using a combined minimum error energy and minimum code length criterion is applied to guarantee good coding performance comparable to many advanced lossless audio coders. The complexity of decoding is relatively small.

REFERENCES

[1] ISO/IEC 11172-3:1993,"Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mbit/s - Part 3:Audio"

[2] ISO/IEC 13818-7:2006 "Information technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding (AAC)".

[3] T. Liebchen. Tech.Univ. Berlin, Berlin,Germany, LPAC, version 0.99h,
http://www-ft.ee.tu-erlin.de/~liebchen/lpac.html.

[4] Eric Knapen, Derk Reefman, Erwin Janssen, and Fons Bruekers, “Lossless compression of one-bit audio”, JAES, vol. 52, no. 2, February 2004.

[5] M. Gerzon et al., “The MLP lossless compression system,” in Proc. AES 17th Int. Conf., Florence, Italy, pp. 61-75, Sept. 1999.

[6] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP): high-quality speech at very low bit rates," in Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 10, pp. 937-940, 1985.

[7] P. Cummiskey, N. S. Jayant, and J. L. Flanagan, "Adaptive quantilation in differential PCM coding of speech," Bell Syst. Tech. J., vol. 52, pp. 1105-1118, Sept. 1973.

[8] ITU-T Recommendation G.726, 40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM), Geneva, Switzerland, 1989.

[9] “FLAC – Free Lossless Audio Codec,” http://flac.sourceforge.net.

[10] “MPEG-4 Audio Lossless Coding (ALS)”
http://www.nue.tu-berlin.de/menue/forschung/projekte/parameter/en/

[11] “Verification Report on MPEG-4 ALS” 74th MPEG Meeting in Nice, France, October 2005.

[12] T. Liebchen, T. Moriya, N. Harada, Y. Kamamoto, and Y. A. Reznik, “The MPEG-4 Audio Lossless Coding Standard - Technology and Applications” 119th AES Convention, New York, NY, USA, Oct. 7-10. 2005, no. 6589, pp. 1-14.