

# ULTRA-LOW LATENCY AUDIO CODING BASED ON DPCM AND BLOCK COMPANDING

*Gediminas Simkus, Martin Holters, Udo Zölzer*

Helmut Schmidt University — University of the Federal Armed Forces  
Department of Signal Processing and Communications  
Holstenhofweg 85, Hamburg, Germany

## ABSTRACT

A low delay audio coding scheme with good perceptual audio quality for a desired limited bit rate is presented. The proposed audio coding scheme is based on differential pulse code modulation (DPCM) and block companded (BC) quantization. Prediction is realized as a FIR filter in lattice structure. DPCM performs in feedback manner, therefore no transmission of prediction filter coefficients is needed. The incorporation of BC quantization in the DPCM relies on a prediction error recalculation scheme. The use of BC quantization in the DPCM allows to accurately follow the prediction error signal. This improves the perceptual audio quality significantly compared to a plain DPCM with an adaptive quantizer. An algorithmic delay below a half millisecond and an overhead of less than a half bit per sample is introduced due to the short fixed block length of the BC quantizer. Therefore, a real time bidirectional audio application is achievable.

**Index Terms**— Audio coding, linear predictive coding, block compander

## 1. INTRODUCTION

Real-time bidirectional digital communication of a performer with a wireless microphone and a wireless headphone as a monitor requires minimal latency. For the real-time audio transmission the algorithmic latency should be less than 5 ms [1]. Besides the low latency, the live microphone-headphone scenario requires a good audio quality for the desired bit rate. The state-of-the-art lossy audio coding schemes such as AAC-ELD codec family provide a nearly transparent audio quality and high signal compression. For this reason, these codecs found wide usage in applications for the storage and broadcast of an audio signal. The algorithmic properties of the AAC-ELD codec lead to a delay as low as 15 ms. This algorithmic delay is introduced during encoding and decoding due to the transform-based block-wise processing in the frequency domain. The audio codecs which use large block lengths are in general not feasible for latency critical audio applications.

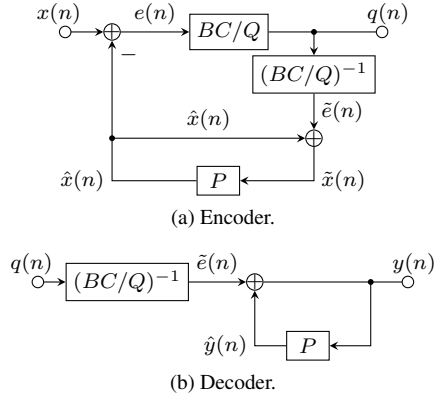
In this work we present a predictive coder which introduces a very low algorithmic delay only during the encoding process and provides a good perceptual audio quality for a desired bit rate.

## 2. THE PROPOSED SYSTEM

The proposed audio coding scheme relies on the well-known base ADPCM coding scheme [2]. The proposed system can be seen as a replacement of the adaptive quantizer in the ADPCM by block companded quantization [3]. The block compander divides a signal into short blocks. A single block consists of less than twenty samples and therefore introduces a short algorithmic delay to the coding scheme. The blocks are normalized and the normalization factors are transmitted to the decoder as side information. This overhead is small compared to the code word length per sample. Besides the mentioned drawbacks of the block companded quantization, the advantage compared to sample-by-sample adaptive quantization is that no clipping of the decorrelated signal appears. The proposed system improves the perceptual audio quality by coding transient signals and provides audio quality near to those coding schemes which use noise shaping techniques [4].

### 2.1. DPCM

The base structure of the standard DPCM coding scheme is depicted in Fig. 1. This scheme corresponds to the base codecs from [2, 4]. Encoder and decoder contain two main blocks, the prediction filter  $P$  and the scalar quantizer  $Q$ . The prediction filter operates in feedback manner. First, for the given current input  $x(n)$  to the encoder, a current prediction  $\hat{x}(n)$  is subtracted. The resulting residual signal is called prediction error  $e(n) = x(n) - \hat{x}(n)$ . The reconstructed prediction error  $\tilde{e}(n) = Q^{-1}(q(n))$  is added up with the current prediction  $\hat{x}(n)$  in the encoder and  $\hat{y}(n)$  in the decoder for the reconstruction of the current input signal, where  $Q^{-1}(\cdot)$  denotes the dequantization operation. This reconstructed signal  $\tilde{x}(n) = \hat{x}(n) + \tilde{e}(n)$  and  $y(n)$  is fed back to the encoder and decoder prediction filter, respectively. Obviously, the



**Fig. 1.** Structure of the DPCM Codec.

reconstructed signals in encoder and in decoder are equal  $\hat{x}(n) = y(n)$ , if no transmission error occurs. As the predictor in encoder and in decoder operates with the same signal values, it is not necessary to transmit the filter coefficients.

A FIR filter in lattice structure [5] is used for the prediction calculation. The block diagram of a  $p$ th-order prediction error filter in lattice structure is shown in Fig. 2. The signals  $f_m(n)$  and  $b_m(n)$ , where  $m = 0, \dots, p$  and  $p$  denotes prediction order, are used to compute the prediction error  $e(n)$ . The signals  $f_m(n)$  and  $b_m(n)$  at lattice stage  $m$  are recursively obtained by

$$f_m(n) = f_{m-1}(n) - k_m b_{m-1}(n-1) \quad (1)$$

$$b_m(n) = b_{m-1}(n-1) - k_m f_{m-1}(n), \quad (2)$$

where the output  $f_p(n)$  is the prediction error for the input  $f_0(n) = b_0(n) = \tilde{x}(n)$  of the prediction error filter.

The gradient adaptive lattice (GAL) algorithm [6] is applied to obtain reflection coefficients  $k_m(n)$ . The coefficients are updated iteratively according to

$$k_m(n+1) = k_m(n) + \mu_m(n) \cdot (f_m(n)b_{m-1}(n-1) + b_m(n)f_{m-1}(n)). \quad (3)$$

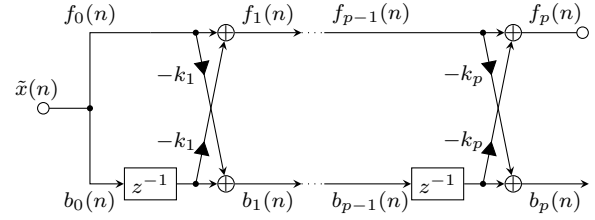
The gradient weights  $\mu_m(n)$  are calculated for every lattice stage by normalizing the base gradient weight  $\tilde{\mu}$  by the energy of signals  $f_{m-1}(n)$  and  $b_{m-1}(n)$  from the previous lattice stage [7]. The gradient weights are given by

$$\mu_m(n) = \frac{\tilde{\mu}}{\sigma_m^2(n) + \sigma_{min}^2} \quad (4)$$

$$\sigma_m^2(n) = (1 - \tilde{\mu})\sigma_m^2(n-1) + \tilde{\mu}(f_{m-1}^2(n) + b_{m-1}^2(n)), \quad (5)$$

where  $\sigma_{min}^2$  is a small constant to avoid division by zero.

The lattice filter stability can be guaranteed by limiting the reflection coefficients to  $|k_m(n)| < 1$ . The GAL algorithm is chosen due to its low computational complexity.



**Fig. 2.** Prediction error filter in the lattice structure.

## 2.2. Block companding

The design of the block companded quantization is presented. First, the signal  $x(n)$  which has to be quantized is divided into blocks of fixed length  $M$ . The introduced algorithmic delay is proportional to the chosen block length  $M$ . Next, the absolute maximal value  $x_{max}(k)$  is calculated for every block, where  $k$  is the current block number. The values in the block  $k$  are normalized by the scaling factor  $x_{max}(k)$ . These normalized values  $q(i) = \frac{x(i)}{x_{max}(k)}$ , where  $i = n - (M - 1), n - (M - 2), \dots, n$  is a time index in the current block, are fed to the quantizer. As the values in the block are normalized, no clipping occurs during the quantization process, if the scaling factor  $x_{max}(k)$  stays unaltered.

For the reconstruction of the quantized signal in the block  $k$ , the scaling factor  $x_{max}(k)$  has to be transmitted to the decoder as side information. The range of maximal absolute values in the block is limited by  $x_{max_{min}} \leq x_{max}(k) \leq 1$ , where  $x_{max_{min}}$  is the minimal scaling factor among all blocks. To prevent big distortions during the quantization process of small valued scaling factors, the same range is logarithmically represented as  $l_{min} \leq 20 \ln(x_{max}(k)) \leq 0$ , where  $l_{min} = 20 \ln(x_{max_{min}})$ . This inequality defines a unipolar quantizer for the coding of  $x_{max_{ln}}(k)$ . To avoid clipping of the values inside the block  $k$ , a new scaling factor  $x'_{max}(k)$  has to be chosen so that

$$x'_{max}(k) \gtrsim \tilde{x}_{max}(k) \quad \text{with} \quad \tilde{x}_{max}(k) = Q^{-1} \left( Q(x_{max}(k)) \right), \quad (6)$$

where  $Q(\cdot)$  and  $Q^{-1}(\cdot)$  are the quantization and the reconstruction operation, respectively.

The overhead per sample is inversely proportional to the block length  $M$ . For example,  $M = 13$  and a word length of  $w_{x_{max}} = 6$  for the bit representation of the scaling factor  $x_{max_{ln}}(k)$  would lead to an overhead of 0.46 bit/sample. Setting the block length  $M$  is deciding between the amount of algorithmic delay, the overhead size per sample and, as shown in the results section 3, the audio quality.

## 2.3. Block companded DPCM

A block companding technique is often applied in sub-band coding schemes [8]. The straight replacement of the fixed

quantizer  $Q$  in the encoder of the DPCM coding scheme in Fig. 1 by the block companded  $BC$  quantizer is impossible due to the sample-wise quantization and reconstruction of prediction error  $\tilde{e}(n)$  which is fed back to the predictor  $P$  to calculate successive predictions. To compute the scaling factor  $x_{max}(k)$ , all  $M$  prediction error samples of the block have to be present. This makes sample-wise encoding unfeasible.

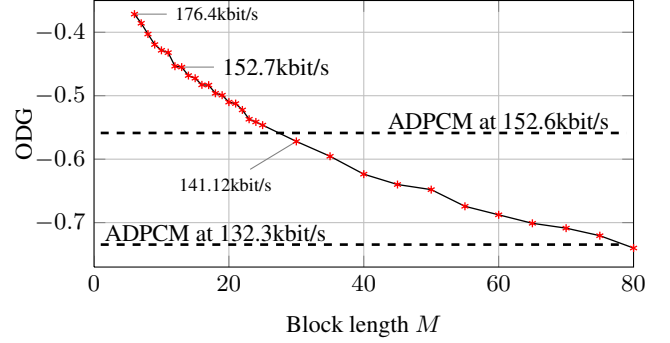
A method to calculate the scaling factor is presented in [9]. Here we present another two step procedure for the  $x_{max}(k)$  calculation, where  $M$  is the block length:

1. Calculate prediction error  $e(j)$  from block start to the block end, where  $j = 1, 2, \dots, M$  is the position index in the block.
  - If  $|e(j)| > x_{max}(k)$  and the number of recalculations limit in the block is not reached  $recalc\_cnt < recalc\_lim$ , then update  $x_{max}(k) = |e(j)|$  and note  $xMax\_changed = 1$  this change of  $x_{max}(k)$  in the block. Proceed till block end is reached.
2. If the block end is reached, i.e. the  $M$ th prediction error is calculated:
  - Repeat step 1 if  $xMax\_changed == 1$  and  $recalc\_cnt \leq recalc\_lim$ . Increase the recalculation counter  $recalc\_cnt = recalc\_cnt + 1$  and reset  $xMax\_changed = 0$ .
  - The block scaling factor  $x_{max}(k)$  is calculated if  $xMax\_changed == 0$ .

After the scaling factor is calculated, first transfer  $x_{max}(k)$  to the decoder and then transfer sample-wise the block samples. Since for the decoder  $x_{max}(k)$  is present, the signal reconstruction is possible sample-wise. Therefore no further algorithmic delay is introduced during the decoding process. To ensure that the same  $x_{max}(k)$  for encoding and decoding is used, the scaling factor  $x_{max}(k)$  in the two step procedure has to be quantized and reconstructed as described in the block compander design section 2.2.

### 3. EVALUATION RESULTS

For the evaluation of the proposed block companded DPCM (BCDPCM) coding scheme, all tracks from SQAM CD [10] are used. The SQAM audio mono excerpts (starting from 0.5 s and 10 s long) with sampling frequency 44.1 kHz are coded using the word length of 3 bit/sample and 6 bit/ $x_{max}(k)$  for payload and scaling factor quantization, respectively. The BCDPCM test parameters are set as follows: the predictor order  $p = 32$ , the base gradient weight  $\tilde{\mu} = 2.08112 \cdot 10^{-3}$  and the minimal energy of lattice stage error signals  $\sigma_{min}^2 = 1.46494 \cdot 10^{-5}$ . The payload quantizer is an uniform asymmetric quantizer. The lower boundary of the uniform unipolar quantizer for the coding of  $x_{max}(k)$  is set to  $l_{min} = -300$ . A noise free channel is used. The audio quality of the



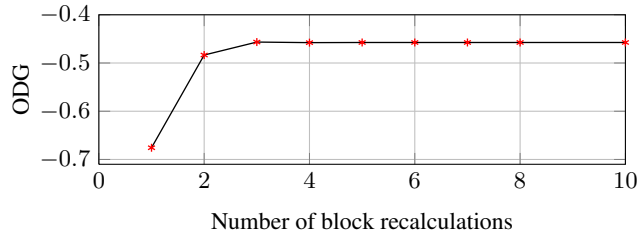
**Fig. 3.** BCDPCM mean perceptual audio quality versus block length  $M$ . Stars mark simulation points starting from  $M = 6$ .

coded test signals are compared to the reference signals based on the ITU-R BS.1387-1 (PEAQ) method [11]. The method classifies the perceptual audio quality by objective difference grades (ODG) on a scale from -4 (very annoying) to 0 (imperceptible).

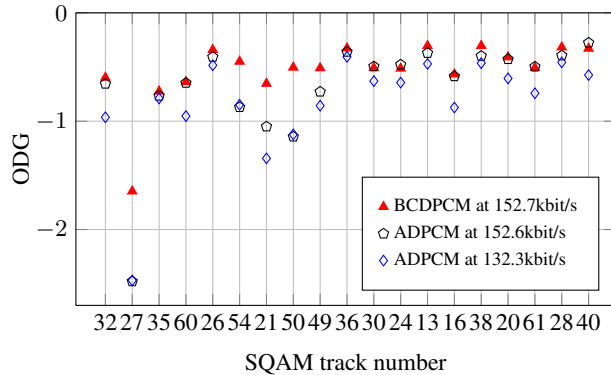
The results of the evaluation for changing block length  $M$  but non-limiting number of block recalculations for the scaling factor calculation are shown in Fig. 3. The first evaluation point with block length  $M = 6$  leads to the best quality and to the shortest delay, but the resulting overhead 1 bit/sample is too high. The mean audio quality over all tracks decreases approximately in a linear manner for increasing block length  $M$ . The dotted lines in Fig. 3 refer to the mean ODG if plain ADPCM with the wordlength of 3 bit/sample and 3.4594 bit/sample is used. These wordlengths correspond to 8 and 11 quantizer levels. The Fig. 3 also clarifies that BCDPCM yields better audio quality compared to ADPCM with a higher number of quantization levels. The block length parameter  $M = 13$  is used for the following evaluation case due to the small delay of 0.29 ms, the measured good mean quality of  $-0.4577$  and a reasonably small overhead per sample of 0.46 bit/sample.

The mean audio quality dependency for different block recalculations is shown in Fig. 4. A significant mean quality improvement is achievable by increasing the block recalculation limit from 1 to 2 and 3. Based on these results, the block recalculation is set to 3 and the audio quality is further considered for the selected SQAM tracks.

The audio quality comparison of BCDPCM and plain ADPCM at nearly identical bitrates is shown in the Fig. 5. Also the results of an optimal ADPCM [12] with 3 bit/sample are marked in Fig. 5. The signals coded by BCDPCM and showing an improvement over both ADPCM at different bitrates clarifies that a signal clipping is avoided if the proposed system is used. The example 27 Castanets clearly shows the direct benefit of the proposed block companded coding scheme. Here the perceptual quality is improved as the short transient signal periods produce transient-like



**Fig. 4.** BCDPCM mean perceptual audio quality versus block recalculation number. Block length  $M = 13$ .



**Fig. 5.** Comparison of BCDPCM and plain ADPCM at different bitrates.

prediction error which are clipped using ADPCM scheme, therefore ODG degrades. The examples, where BCDPCM and ADPCM operate at the same bitrates and yield a higher ODG measure compared to the ADPCM with the wordlength of 3 bit/sample, show that a quality improvement is possible if a higher quantization resolution or a more accurate envelope estimation of the residual signal is used. The examples with nearly the same ODG measure for all the three systems can be explained by a slowly decaying prediction error. In this case block companded and adaptive quantizers are able to accurately follow the envelope of a residual signal.

#### 4. CONCLUSIONS

We have proposed a coding scheme BCDPCM which incorporates block companded quantization with differential pulse code modulation. The proposed system introduces a small delay to the coding scheme due to the scaling factor determination on the encoder side and also adds an overhead. The determination of the scaling factor leads to a more complex coding scheme, but the proposed BCDPCM provides for transient signals significantly better audio quality compared to the plain ADPCM coding scheme. We believe that further sound quality improvements are possible by adding quantization noise shaping techniques to the proposed BCDPCM coding scheme and by optimizing coding parameters.

#### 5. REFERENCES

- [1] Aki Härmä and Unto K. Laine, "Warped low-delay CELP for wideband audio coding," in *Audio Engineering Society Conference: High-Quality Audio Coding*, Aug. 1999.
- [2] D. Cohn and J. Melsa, "The residual encoder – an improved ADPCM system for speech digitization," Sep. 1975, vol. 23, pp. 935–941.
- [3] K. Niwa, T. Araseki, and A. Tomozawa, "A new channel bank with block companding," *Communications, IEEE Transactions on*, vol. 30, no. 4, pp. 574–580, Apr. 1982.
- [4] Martin Holters and Udo Zölzer, "Delay-free lossy audio coding using shelving pre- and post-filters," in *Acoustics, Speech and Signal Processing. ICASSP 2008. IEEE International Conference on*, Apr. 2008, pp. 209–212.
- [5] R. Reininger and J. Gibson, "Backward adaptive lattice and transversal predictors for ADPCM," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '84.*, Mar. 1984, vol. 9, pp. 429–432.
- [6] L. Griffiths, "A continuously-adaptive filter implemented as a lattice structure," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '77.*, May 1977, vol. 2, pp. 683–686.
- [7] C. Gibson and S. Haykin, "Learning characteristics of adaptive lattice filtering algorithms," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 6, pp. 681–691, Dec. 1980.
- [8] D. Esteban and C. Galand, "Application of quadrature mirror filters to split band voice coding schemes," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '77.*, May 1977, vol. 2, pp. 191–195.
- [9] Gediminas Simkus, Martin Holters, and Udo Zölzer, "Ultra-low delay lossy audio coding using DPCM and block companded quantization," in *Proc. 14th Australian Communications Theory Workshop*, Feb 2013.
- [10] EBU Tech 3253, "Sound quality assessment material: Recordings for subjective tests," Apr. 1988.
- [11] ITU-R BS.1387-1, "Method for objective measurements of perceived audio quality," Nov. 2001.
- [12] Martin Holters, Christian R. Helmrich, and Udo Zölzer, "Delay-free audio coding based on ADPCM and error feedback," in *Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Sep. 2008.