ERROR ROBUST DELAY-FREE LOSSY AUDIO CODING BASED ON ADPCM

Gediminas Simkus, Martin Holters, Udo Zölzer

Dept. of Signal Processing and Communications,
Helmut Schmidt University — University of the Federal Armed Forces
Hamburg, Germany
gediminas.simkus|martin.holters|udo.zoelzer@hsu-hh.de

ABSTRACT

We consider the problem of transmission errors in the well known adaptive differential pulse code modulation (ADPCM) system. A single transmission error destabilizes the reconstruction process at the decoder side in the ADPCM coding scheme if a non-leaky algorithm is used. We propose a delay-free and fixed rate of 3 bit/sample audio source coding scheme based on a robust prediction. The prediction of the backward ADPCM coding scheme is attained as a FIR filter in lattice structure. The prediction filter is derived as a reconstructed-signal-driven (RSD) or a prediction-error-driven (PED) lattice filter. A technique for an error robust RSD prediction is presented. This technique is employed in a robust audio coding scheme without use of any additional overhead. The proposed modified RSD-ADPCM is compared to the PED-ADPCM coding scheme by means of the objective audio quality. The proposed system yields good objective audio quality in the noise-free channels and provides robustness in the presence of transmission errors.

1. INTRODUCTION

Real-time bidirectional digital communication of a performer with a wireless microphone and a wireless headphone as a monitor requires minimal latency and good audio quality for restricted channel capacity. Furthermore, robustness against channel transmission errors must be guaranteed, if such a wireless application is applied. A typical scenario where the minimal latency requirement applies, is the music stage. A vocalist may lose the beat if his voice that comes from the monitors is delayed. In the same manner a musician is unable to keep the rhythm if his performance is delayed.

Well known audio coding schemes such as MPEG-Layer 3 (MP3) and MPEG-4 (AAC) provide a nearly transparent audio quality and high signal compression. These coding schemes are based on block-transformation from time to frequency domain. Such audio coding schemes which use large transformation blocks introduce large algorithmic latency. Therefore, the latency of the MP3 coding scheme can exceed 100 ms [1]. As a consequence, a coding scheme with a long block processing is not suitable for a live audio application. The algorithmic latency of such an application should be less than 5 ms [2]. Further developments of the AAC coding scheme allow to reduce the algorithmic latency down to 15 ms [3].

Audio coding approaches, discussed in [4, 5, 6], concentrate on audio coding without any algorithmic latency and good audio quality, if an error-free transmission channel is used. These audio source coding schemes are based on ADPCM [7, 8]. The goal of such an audio coding scheme is the digital real-time wireless audio application. However, transmission errors occur in a wireless application. For this reason, we examine the effect of transmission channel errors to the base ADPCM in this work. Especially, we discuss the predictor, the main part of ADPCM system. The prediction is calculated by using a prediction filter in lattice structure [9]. In this work we concentrate on two classes of lattice prediction filters [10] used in the ADPCM coding schemes: reconstructed-signal-driven (RSD) and prediction-error-driven (PED). Furthermore, we provide a robust RSD-predictor for transmission error robust ADPCM coding. Therefore, the aim of this work is to find a robust lattice prediction filter, which is suitable for a delay-free ADPCM audio coding scheme, if an error-free or noisy transmission channel is used. Besides that, the robust audio coding scheme has to guarantee good audio quality for a desired bandwidth.

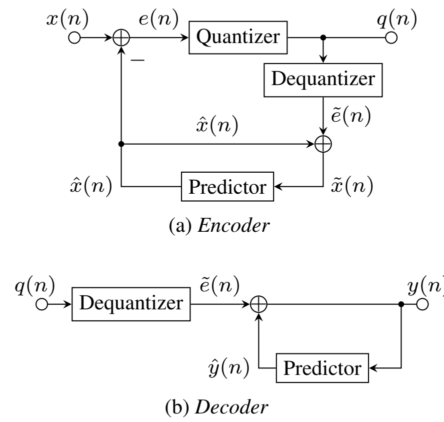
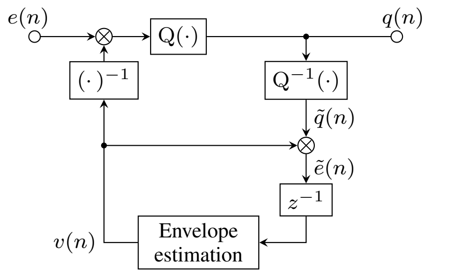


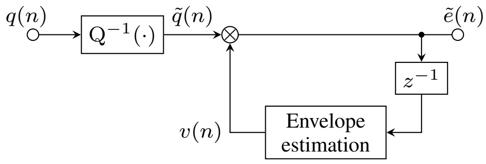
Figure 1: *Structure of the base ADPCM coding scheme.*

2. THE BASE ADPCM CODEC

The central issue of a well-known ADPCM coding scheme is the decorrelation of the input signal. The base ADPCM coding scheme depicted in Fig. 1 corresponds to the base codecs from [5, 7, 8]. Two main methods, prediction filter and adaptive quantizer, are employed by the ADPCM for audio signal encoding and decoding. The prediction $\hat{x}(n)$ at the encoder and at the decoder is generated from previous prediction filter inputs. The prediction filter operates in feed-back manner, consequently, it is not necessary to transmit any side information. The current prediction $\hat{x}(n)$ is subtracted from current input $x(n)$. As a result, the prediction error $e(n) = x(n) - \hat{x}(n)$ is quantized by a backward adaptive



(a) *Adaptive quantization*



(b) *Adaptive dequantization*

Figure 2: *Structure of the backward adaptive quantizer.*

quantizer to yield the desired bit rate. The backward adaptive quantizer depicted in Fig. 2 is similar to the approaches presented in [6, 11]. The prediction error $e(n)$ is normalized at the encoder by its estimated envelope $v(n)$ and then quantized by a quantization operation $\mathrm{Q}(\cdot)$ to $q(n) = \mathrm{Q}\left(\frac{e(n)}{v(n)}\right)$. The resulting quantized normalized prediction error $q(n)$ is transmitted to the decoder. The reconstructed prediction error $\tilde{e}(n) = \mathrm{Q}^{-1}\left(q(n)\right) \cdot v(n)$ is added up with the current prediction at the encoder and at the decoder, where $\mathrm{Q}^{-1}(\cdot)$ denotes the dequantization operation. The reconstruction of the original signal is done at the encoder by $\tilde{x}(n) = \hat{x}(n) + \tilde{e}(n)$ and at the decoder by $y(n) = \hat{y}(n) + \tilde{e}(n)$. The reconstructed signal $\tilde{x}(n)$ and $y(n)$ is fed to the encoder and decoder prediction filter, respectively.

The envelope or the scaling factor $v(n)$ is estimated by calculating the instantaneous power $v^2(n)$ of the reconstructed prediction error signal $\tilde{e}(n)$ as follows:

$$v^2(n) = \left(1 - \lambda\right) \cdot v^{2\beta}(n-1) + \lambda \cdot \tilde{e}^2(n-1), \qquad (1)$$

$$\lambda = \begin{cases} \lambda_{AT} & \text{if } v^2(n-1) < \tilde{e}^2(n-1) \\ \lambda_{RT} & \text{otherwise} \end{cases} \quad \text{with } \lambda_{AT} > \lambda_{RT}.$$

The parameter λ controls how fast the estimate $v^2(n)$ follows the signal. The attack-time parameter λ_{AT} is taken if the power of the prediction error signal $\tilde{e}(n)$ increases, otherwise the release-time parameter λ_{RT} is used.

To avoid division by zero during the signal power normalization $q(n) = \mathrm{Q}\left(\frac{e(n)}{v(n)}\right)$, the estimated envelope $v(n)$ is bounded to $0 < v_{min} \leq v(n)$.

The backward adaptive quantizer envelope estimation robustness against transmission errors is ensured in Eq. (1) by adding the damping parameter β. Therefore, the impact of previously calculated scaling factor $v(n-1)$ to the current estimate $v(n)$ is damped. The parameter β chosen from $0 < \beta < 1$ guarantees that the estimate at the decoder after certain time becomes the same as at the encoder, if a channel transmission error occurs.

For instance, $\lambda_{AT} = 0.85$, $\lambda_{RT} = 0.1$, $v_{min} = 2^{-10}$ and $\beta = 1 - 2^{-9}$ are appropriate adaptive quantizer parameters to use in the base ADPCM coding scheme.

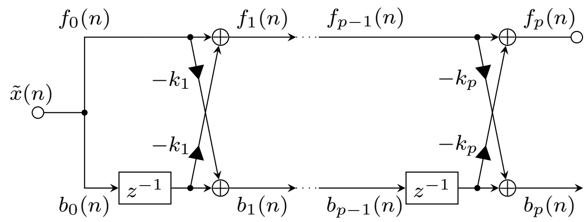


Figure 3: *Prediction error filter in the lattice structure.*

2.1. RSD-ADPCM

The prediction applied to the ADPCM coding scheme shown in Fig. 1 is calculated by a FIR filter in lattice structure [12, 13]. The block diagram of a pth-order prediction error filter in lattice structure is illustrated in Fig. 3. The forward $f_m(n)$ and backward $b_m(n)$ prediction errors, where $m = 0, \ldots, p$ and p denotes prediction order, are used to compute the desired prediction $\hat{x}(n)$. The signals $f_m(n)$ and $b_m(n)$ at lattice stage m are recursively obtained by

$$f_m(n) = f_{m-1}(n) - k_m b_{m-1}(n-1) \qquad (2)$$

$$b_m(n) = b_{m-1}(n-1) - k_m f_{m-1}(n). \qquad (3)$$

The filter output $f_p(n) = e(n)$ is the desired prediction error for the filter input $f_0(n) = b_0(n) = \tilde{x}(n)$. The desired prediction is calculated by $\hat{x}(n) = \tilde{x}(n) - f_p(n)$ or by

$$\hat{x}(n) = \sum_{m=1}^{p} k_m(n) b_{m-1}(n-1). \qquad (4)$$

The gradient adaptive lattice (GAL) algorithm [14, 15] is applied to calculate lattice filter reflection coefficients $k_m(n)$. GAL method updates coefficients $k_m(n)$ iteratively by

$$\begin{aligned} k_m(n+1) &= k_m(n) + \\ &\mu_m(n) \cdot \big(f_m(n) b_{m-1}(n-1) + b_m(n) f_{m-1}(n)\big). \end{aligned} \qquad (5)$$

The gradient weights $\mu_m(n)$ [6, 16] are calculated for every lattice stage by normalizing the base gradient weight $\tilde{\mu}$ by the energy of the previous lattice stage inputs which are the forward $f_{m-1}(n)$ and backward $b_{m-1}(n-1)$ prediction errors. The gradient weights $\mu_m(n)$ are given then by

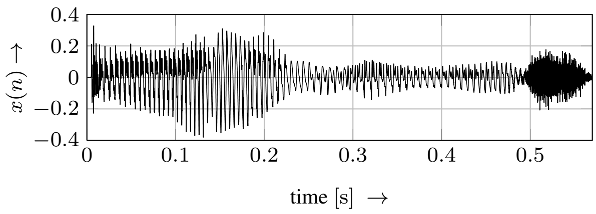
$$\mu_m(n) = \frac{\tilde{\mu}}{\sigma_m^2(n) + \sigma_{min}^2} \qquad (6)$$

$$\begin{aligned} \sigma_m^2(n) &= (1 - \tilde{\mu})\sigma_m^2(n-1) + \\ &\tilde{\mu}\big(f_{m-1}^2(n) + b_{m-1}^2(n-1)\big), \end{aligned} \qquad (7)$$

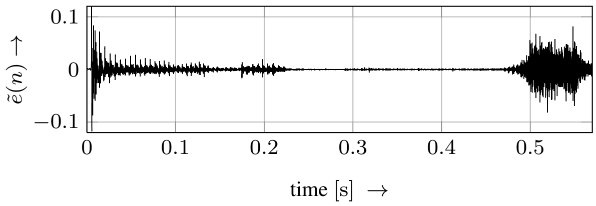
where σ_{min}^2 is a small constant to avoid division by zero.

The important property of the lattice filter structure is the minimum phase property, so that the filter stability can be guaranteed by limiting the filter coefficients to $|k_m(n)| < 1$.

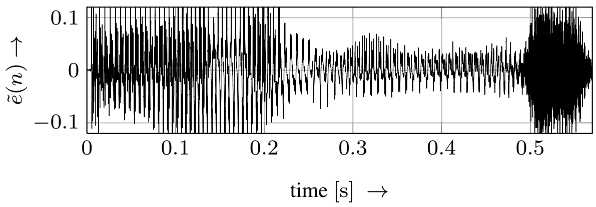
The input to the lattice predictor is the reconstructed signal $\tilde{x}(n)$ and the filter coefficient adaptation is also employed by $\tilde{x}(n)$. Therefore, such a filter is entitled in this paper as reconstructed-signal-driven (RSD) lattice filter. The exploitation of RSD lattice filter in the base ADPCM coding scheme is named as RSD-ADPCM.



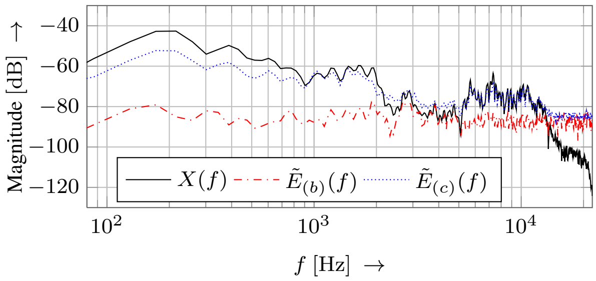
(a) *Input signal*



(b) *Base RSD-ADPCM*



(c) *Leaky GAL ($\alpha = 0.99$) in RSD-ADPCM*



(d) *Power spectra of signals in Fig. (a), (b) and (c)*

Figure 4: *The word "animal" from the SQAM track 49 as the signal input and the corresponding prediction error signals with the power spectra, using different prediction filter adaptation algorithms.*

As shown in Fig. 4b, the RSD-ADPCM system exhibits high correlation reduction in the middle part of signal compared to the original signal in Fig. 4a. The high signal correlation reduction of a stationary signal as the input to the RSD-ADPCM encoder is achieved by the jointly action of the RSD lattice filter and the backward adaptive quantizer. A good prediction to a current encoder input is produced by the RSD lattice filter, which employs the iterative reflection coefficient adaptation from the reconstructed signal. Due to the signal correlation reduction, a smaller word length may be used to transmit a small powered prediction error compared to a high powered original signal. Therefore, the bit rate of the RSD-ADPCM may be reduced by using the adaptive quantizer to the desired rate of 3 bit/sample. As shown in Section 3, a good

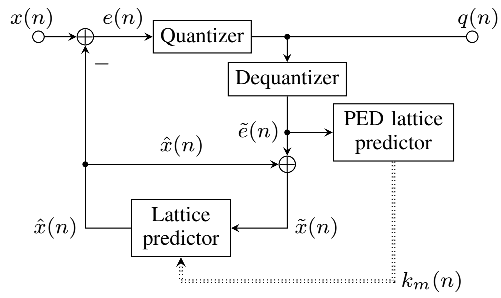


Figure 5: *Structure of the PED-ADPCM encoder.*

audio quality is guaranteed, if an error-free channel is used.

The encoder and the decoder of the RSD-ADPCM system operate with identical coefficients, if no transmission errors occur. A transmission error on the prediction error signal $\tilde{e}(n)$ affects the prediction calculation in the RSD lattice filter at the decoder side $\hat{y}(n)$. Consequently, the decoder produces an erroneous output. As the erroneous reconstructed signal is fed to the decoder RSD lattice predictor, the GAL algorithm calculates erroneous reflection coefficients. Thus, the decoder predictor operates with different values than the encoder predictor. The erroneous prediction calculation at the decoder starts to propagate in the successive reconstructed signal samples. All reconstructed samples are erroneous after a single channel error occurs. The problem of the deviated values of the decoder predictor from the encoder predictor is referred to as mistracking. The mistracking of the decoder predictor from the encoder predictor is produced essentially by the iterative coefficient adaptation of the RSD lattice filter. To avoid the mistracking of the decoder from the encoder after a transmission error occurs, either a certain channel coding with error detection and correction, or a transmission of the encoder filter states to the decoder, or a change of the filter adaptation algorithm is necessary. The first two mentioned methods lead to an overhead in the bit rate and possibly to an algorithmic delay. Thus, a GAL algorithm modification is considered. If a leakage factor $\alpha < 1$ is added to the reflection coefficient update method in Eq. (5), then the reflection coefficients are iteratively calculated at the encoder and decoder by

$$\begin{aligned} k_m(n+1) &= \alpha \cdot k_m(n) + \\ &\mu_m(n) \cdot \big(f_m(n) b_{m-1}(n-1) + b_m(n) f_{m-1}(n)\big). \end{aligned} \tag{8}$$

Therefore, decoder lattice filter values converge in long term to the encoder values, if a transmission error occurs. The leakage on the reflection coefficients $k_m(n)$ affects the increase in the prediction error power. This power increase is clearly seen in the waveform Fig. 4c and in the power spectra in Fig. 4d, if $\alpha = 0.99$. The corresponding input signal is shown in Fig. 4a. Consequently, a small word length per sample can not ensure good audio quality, if a leaky GAL in the base RSD-ADPCM is used.

2.2. PED-ADPCM

The authors in [10] examine the mistracking problem at the decoder of the lattice prediction filter and propose a new class of prediction-error-driven (PED) lattice filter. The structure of the PED-ADPCM encoder is depicted in Fig. 5. The PED method is based on the following heuristic approach. The employment of non-optimal reflection coefficients for the prediction error calculation affects the prediction error $\tilde{e}(n)$. Thus, the signal $\tilde{e}(n)$ power increases and $\tilde{e}(n)$ is correlated with the reconstructed signal $\tilde{x}(n)$. Moreover, the reflection coefficients of two correlated signals are also correlated. Therefore, the reflection coefficients, which are calculated in the first lattice filter (PED lattice predictor block in Fig. 5) driven by the prediction error signal $\tilde{e}(n)$, may be used to drive the second lattice predictor. The input of the second lattice prediction filter is the reconstructed signal $\tilde{x}(n)$.

The decorrelation of the prediction error signal $\tilde{e}(n)$ in PED lattice filter is done similarly according to Eq. (2) and (3) by

$$\bar{f}_m(n) = \bar{f}_{m-1}(n) - k_m \bar{b}_{m-1}(n-1) \quad (9)$$

$$\bar{b}_m(n) = \bar{b}_{m-1}(n-1) - k_m \bar{f}_{m-1}(n), \quad (10)$$

with the filter input $\bar{f}_0(n) = \bar{b}_0(n) = \tilde{e}(n)$. The adaptation of the reflection coefficients is done as in Eq. (5), (6) and (7) using signals calculated in Eq. (9) and (10). These residual signals are not used for the calculation of the desired prediction error $e(n)$. The reflection coefficients $k_m(n)$ from the PED lattice filter are directly used for the calculation of the desired prediction $\hat{x}(n)$ in the second lattice filter.

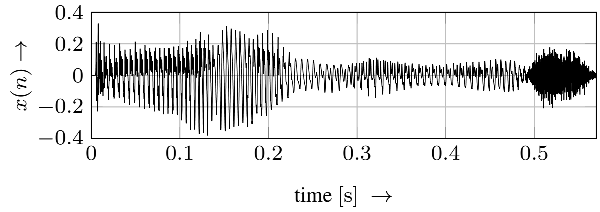
As depicted in Fig. 6b and 6d, the signal power reduction using the PED-ADPCM encoder with the input signal shown in Fig. 6a is evident. In comparison, as depicted in Fig. 4b, the prediction error power of the same input signal coded by the RSD-ADPCM encoder is smaller. Therefore, as will be presented in Section 3, the PED-ADPCM system needs a higher word length to exhibit the same audio quality compared to the RSD-ADPCM coding scheme, if an error-free channel is employed.

The advantage concerning the error robustness of the PED-ADPCM coding scheme is that the PED lattice filter adapts reflection coefficients from the residual signals calculated in lattice stages. The first lattice stage input is the prediction error $\tilde{e}(n)$. If a transmission channel causes an error on $\tilde{e}(n)$, then this error propagates in the reconstructed signal, as shown in Fig. 7b, only for a certain time period. This error propagation is limited, because the signals $\bar{f}_m(n)$ and $\bar{b}_m(n)$ at the decoder side can converge to the same values as at the encoder side, if no further transmission errors appear. Therefore, the convergence of the reflection coefficients $k_m(n)$ at the decoder to the encoder values is also possible. Nevertheless, the same transmission error propagates endless in the non-leaky RSD-ADPCM system, because the reconstructed and corrupted signal $\tilde{x}(n)$ is the input to the first lattice stage.

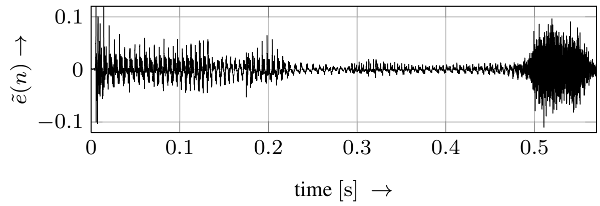
The PED-ADPCM coding scheme is robust against burst errors. A corresponding coding error $e_{cod}(n) = x(n) - y(n)$ caused by a burst error of 0.75 s duration is depicted in Fig. 8b. To ensure robustness against burst errors, a reset of the lattice filter states is done at the decoder side if the reconstructed signal $y(n) \gg 1$. Additionally, the estimated scaling factor $v(n)$ in Eq. (1) of the adaptive quantizer is set to a fixed value $v_{fix} = 10^{-2}$. After receiving several erroneous prediction error indexes, the output of the decoder is muted. As the decoder values start to converge to the encoder values, the decoder output becomes available.

2.3. Proposed error robust RSD-ADPCM

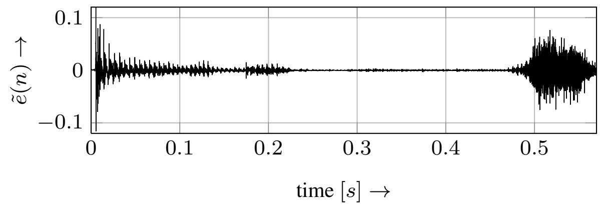
A modification of the discussed RSD-ADPCM for employment in noisy transmissions is presented. The proposed RSD lattice error prediction filter is shown in Fig. 9. As described in Section 2.1, the decay of the coding error is impossible due to the use of a reconstructed signal in the signal coding. Hence, to prevent signal



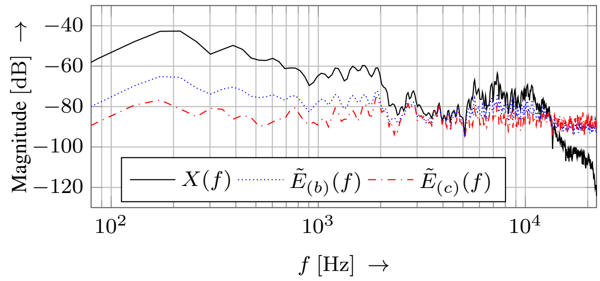
(a) *Input signal*



(b) *PED-ADPCM*



(c) *Proposed RSD-ADPCM*



(d) *Power spectra of (a), (b) and (c)*

Figure 6: *Input signal coded by different ADPCM coding schemes and corresponding power spectra.*

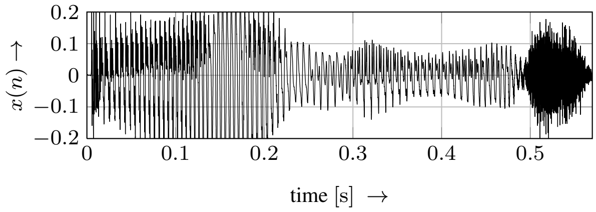
mistracking at the decoder side, two damping parameters are introduced into RSD lattice filtering algorithm. The first damping parameter α is used for leakage of the lattice states $b_{m-1}(n-1)$. Therefore, the forward and backward prediction error signals are calculated by

$$f_m(n) = f_{m-1}(n) - k_m \cdot \alpha b_{m-1}(n-1) \quad (11)$$

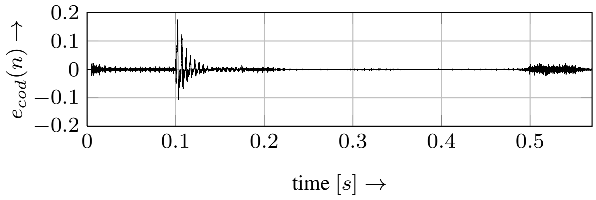
$$b_m(n) = \alpha b_{m-1}(n-1) - k_m \cdot f_{m-1}(n), \quad (12)$$

where the damping parameter is set to $\alpha < 1$.

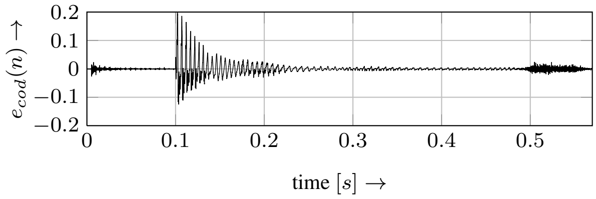
As the reflection coefficients are highly vulnerable to the transmission errors, the damped backward prediction error signal $\alpha b_{m-1}(n-1)$, which is the input to the GAL algorithm, is additionally damped by a second parameter β. The damping parameter is chosen to be $\beta < \alpha$. The second damping parameter is used to reduce the impact of the possible transmission error on



(a) *Input signal*



(b) *PED-ADPCM coding error*



(c) *Proposed RSD-ADPCM coding error*

Figure 7: *Example of the most-significant bit error at 0.1 s. Original input and corresponding coding errors $e_{cod} = x(n) - y(n)$, if PED-ADPCM and proposed RSD-ADPCM are used.*

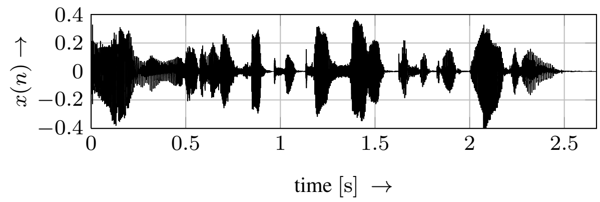
the reflection coefficient adaptation algorithm. Consequently, the update of the reflection coefficients $k_m(n)$ is done similarly to the Eq. (5) and is obtained by

$$k_m(n+1) = k_m(n) + \mu_m(n) \cdot \big(f_m(n) \cdot \alpha\beta b_{m-1}(n-1) + b_m(n) \cdot f_{m-1}(n)\big). \quad (13)$$

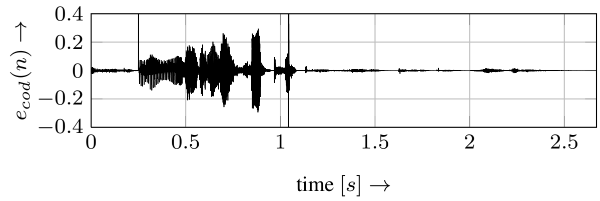
The gradient weights $\mu_m(n)$ are calculated without changes as in Eq. (6) and (7). Therefore, after the reflection coefficients are updated, the prediction for the following input sample may be calculated similarly as in Eq. (4) by

$$\hat{x}(n) = \sum_{m=1}^{p} k_m(n) \cdot \alpha b_{m-1}(n). \quad (14)$$

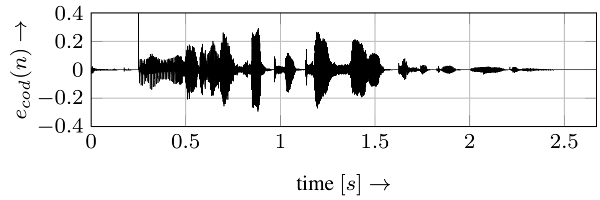
Although the use of damping factors for the lattice prediction calculation affects the power of the desired prediction error, the increase of the prediction error power is minor and the use of short word length is possible. As shown in Fig. 6c and 6d, the proposed RSD-ADPCM coding scheme guarantees high correlation reduction compared to the original signal. The comparison of the power spectra examples coded with the non-error robust RSD-ADPCM (Fig. 4d) and the proposed RSD-ADCM (Fig. 6d) shows similar power. Thus, a similar audio quality is expected of both coding schemes at the same bit rates and in the error-free transmissions. The parameters of the proposed RSD-ADPCM system in the shown examples are chosen as in Table 1.



(a) *Input signal*



(b) *PED-ADPCM coding error*



(c) *Proposed RSD-ADPCM coding error*

Figure 8: *Example of channel transmission burst error starting at 0.25 s and ending at 1 s. Original input and corresponding coding errors, if PED-ADPCM and proposed RSD-ADPCM are used.*

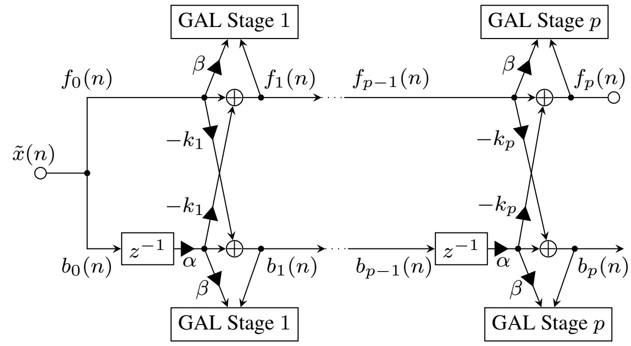


Figure 9: *Channel error robust prediction error filter in the lattice structure with signal outputs for GAL adaptation at every lattice stage.*

The proposed modification for the RSD-ADPCM coding scheme with particular damping parameters ensures that the decoder may converge in long term to the same values as at the encoder side, if the reconstructed signal $y(n)$ is affected by a transmission error. The Fig. 7c illustrates the coding error, if a erroneous transmission channel is used. The decay of the error $d(n) = |y_{\text{noise-free}}(n) - y_{\text{noisy}}(n)|$ caused by a transmission error is considerably longer compared to the PED-ADPCM error, see Fig. 7b. In general, the complete decay of error $d(n)$ takes

Table 1: *Proposed RSD-lattice filter parameters.*

Parameter	Value
α	0.98
β	0.91
σ^2_{min}	$1.8220 \cdot 10^{-10}$
$\tilde{\mu}$	$1.3498 \cdot 10^{-3}$

approximately 8 s, if the input is a sinusoidal signal and the coding parameters are set as in Table 1. The parameters are found experimentally using different power sinusoidal signals. These parameters guarantee that the error $d(n)$ may decay in long term to $d(n) = 10^{-12}$. The error decay to $d(n) = 10^{-4}$ is reached after approximately 1.2 s and after next 6.5 s the error is smaller than 10^{-12}. In comparison, using PED-ADPCM coding scheme the same transmission error decays after approximately 0.3 s and 0.8 s to $d(n) = 10^{-4}$ and $d(n) = 10^{-12}$, respectively. Nevertheless, as shown in Section 3, the perceptual audio quality of the proposed RSD-ADPCM is significantly better compared to the PED-ADPCM coding scheme at the same bit rates, if an error-free channel is used. Additionally, the perceptual audio quality of the proposed robust coding scheme is only slightly worse compared to the non-robust RSD-ADPCM coding scheme.

As shown in Fig. 8c, the decay of the coding error is also possible if a burst error occurs. In general, the coding error decays after a long burst error to $d(n) = 10^{-4}$ and $d(n) = 10^{-12}$ for the sinusoidal input after approximately 3 s and 9 s, respectively. Using the same disturbed sinusoidal signal in the PED-ADCM coding scheme the error $d(n)$ decays to $d(n) = 10^{-4}$ and $d(n) = 10^{-12}$ after 0.3 s and 1 s, respectively.

3. EVALUATION RESULTS

The evaluation of ADPCM coding schemes using PED and RSD lattice filters is considered. The perceptual audio quality is compared between RSD-, PED- and proposed RSD-ADPCM coding schemes, if error-free and erroneous transmission channel is used. All tracks from SQAM CD [17] are used to evaluate the presented audio coding schemes. The SQAM audio mono excerpts (starting from 0.5 s and 10 s long) with sampling frequency 44.1 kHz are coded using word length of 3 bit/sample and 4 bit/sample. All coding schemes use prediction order of $p = 32$. The RSD- and PED-ADPCM parameters are set to the values proposed in [6]. Additionally, the parameters of the proposed RSD-ADPCM coding scheme are set as in Table 1. The audio quality of the coded test signals are compared to the reference signals based on the ITU-R BS.1387-1 (PEAQ) method [18, 19]. The method classifies the perceptual audio quality by objective difference grades (ODG) on a scale from -4 (very annoying impairment) to 0 (imperceptible impairment).

3.1. Audio quality

The mean perceptual audio quality over all SQAM tracks in this paper discussed ADPCM coding schemes is depicted in Fig. 10. To evaluate the perceptual audio quality, it is assumed that the transmission channel is error free. This assumption is performed to compare a non-robust RSD-ADPCM coding scheme with a robust (proposed RSD-ADPCM and PED-ADPCM) audio coding schemes. The comparison in Fig. 10 shows that for both bit rates

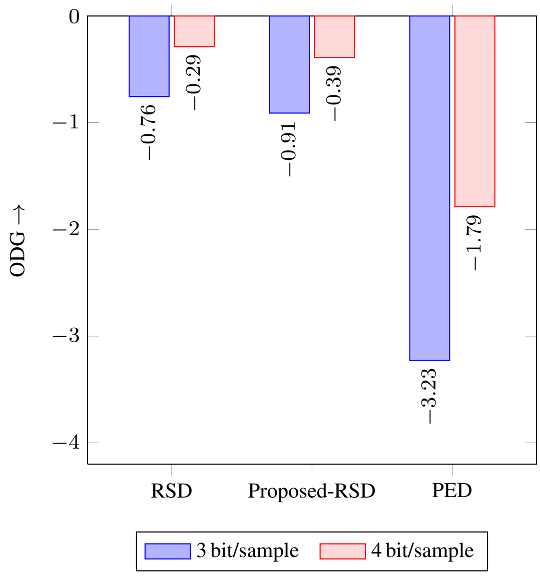


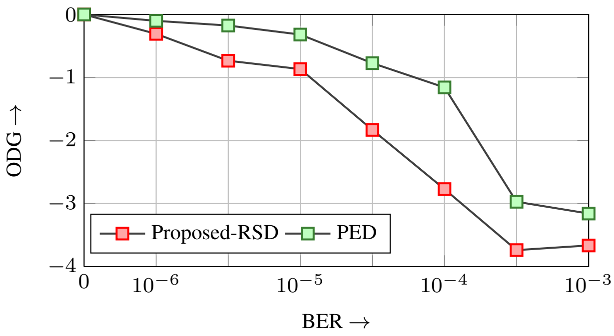
Figure 10: *Mean perceptual audio quality by employing different lattice prediction (RSD, proppposed-RSD, PED) filters in the ADPCM coding scheme.*

the proposed RSD-ADPCM system exhibits similar audio quality by means of ODG compared to the non-robust RSD-ADPCM coding scheme. In contrast, the PED-ADPCM coding scheme with word length of 4 bit/sample underperforms significantly compared to other coding schemes which use smaller bit rates. For this reason, the PED-ADPCM coding scheme with a higher word length is evaluated. Only at 5 bit/sample with the mean audio quality of -0.7085 ODG and of -0.3174 ODG at 6 bit/sample the PED-ADPCM coding scheme exhibits similar ODG as the RSD-ADPCM coding schemes at 3 bit/sample and 4 bit/sample, respectively.

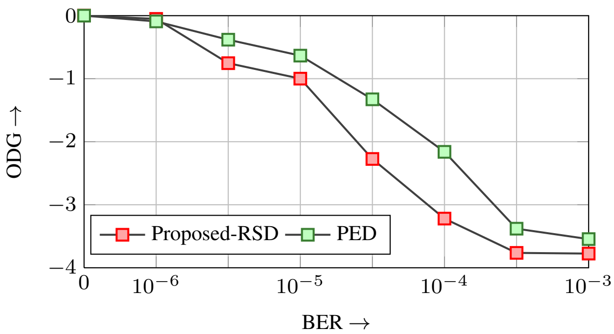
3.2. Robustness

The robustness against transmission errors is compared between PED- and the proposed RSD-ADPCM coding schemes. The PEAQ method is used to compare robustness of each coding scheme. This evaluation is done as follows. First, a signal decoded by the corresponding coding scheme is used as the reference signal $y_{\text{noise-free}}(n)$, where an error-free channel is used. Second, the test signals $y_{\text{noisy}}(n)$ are obtained by decoding with different bit error rates (BER) disturbed signals. Finally, the reference signal is compared to the disturbed test signals. Comparing error-free $y_{\text{noise-free}}(n)$ and noisy signals $y_{\text{noisy}}(n)$ allows to evaluate the impact of bit errors on signals at the corresponding BER. Therefore, the influence of the quantization noise is ignored. The perceptual audio quality of a track depends on which sample in the track a bit error appears. Accordingly, all SQAM tracks at the corresponding BER and different coding schemes are disturbed by the same bit error pattern.

As discussed in Section 2, the transmission error decay using PED-ADPCM coding scheme is faster compared to the proposed RSD-ADPCM. Therefore, a higher ODG is expected, if PED-ADPCM is used under noisy conditions. The robustness comparison of the proposed RSD-ADPCM and PED-ADPCM

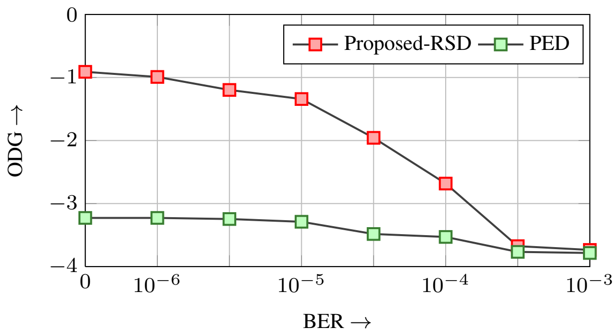


(a) *3 bit/sample*

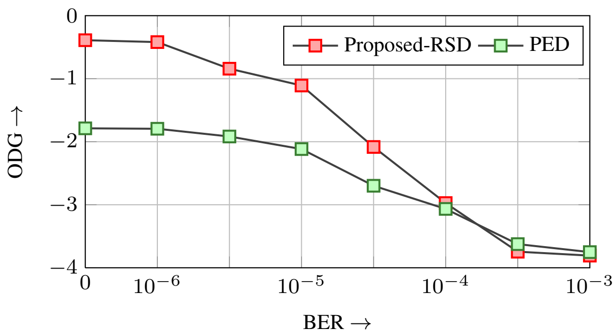


(b) *4 bit/sample*

Figure 11: *Impairment evaluation of the decoded signals* $y_{noisy}(n)$ *at different BERs by employing robust ADPCM coding schemes. The PEAQ algorithm compares the decoded noise-free signal* $y_{noise\text{-}free}(n)$ *with* $y_{noisy}(n)$.



(a) *3 bit/sample*



(b) *4 bit/sample*

Figure 12: *Overall mean perceptual audio quality against BER by employing robust ADPCM coding schemes, if the PEAQ algorithm compares the original* $x(n)$ *with the decoded noisy signal* $y_{noisy}(n)$.

coding schemes at different word lengths is depicted in Fig. 11. As expected, the PED-ADPCM coding scheme exhibits better audio quality for both word lengths and tested BERs. As shown in Fig. 11a, the PED-ADPCM codec yields in the mean perceptible but not annoying signal impairment up to BER $= 10^{-4}$. The same robustness of the proposed-ADPCM may be reached for BER $= 10^{-5}$. The impairment of the coded signals is classified as annoying, if the proposed-ADPCM coding scheme and the bit error pattern of BER $= 10^{-4}$ is used. As denoted in Fig. 11b, increasing the word length to 4 bit/sample affects the robustness performance of both codecs in the decrease of ODGs at all tested BERs. In general, both coding schemes up to BER $= 10^{-5}$ yield similar good robustness against transmission errors. For higher BERs the PED-ADPCM codec shows better error robustness properties. The SQAM tracks coded with both ADPCM methods are classified as very annoying, if BER reaches 10^{-3}.

The final evaluation is done by using the original signal $x(n)$ as the reference signal and the decoded noisy signal $y_{\text{noisy}}(n)$ at the corresponding BER as the test signal. Therefore, the quantization noise is not ignored by the PEAQ algorithm and the overall perceptual impairment of the decoded signal is evaluated. As shown in Fig. 12, the proposed-RSD ADPCM method nearly at all BERs yields higher ODG compared to the PED-ADPCM coding scheme. In general, the proposed-RSD ADPCM guarantees slightly annoying impairment of the decoded signal up to BER of $5 \cdot 10^{-5}$. Nearly the same ODG may be yielded by the PED-ADPCM coding scheme, if the word length is set to 4 bit/sample and the transmission is error-free or up to BER of $5 \cdot 10^{-6}$. The evaluation in Fig. 11 of the PED-ADPCM shows better robustness properties, if the quantization noise impairment is ignored. However, if the overall impairment is evaluated (Fig. 12), the proposed-RSD ADPCM system in the noisy and error-free channels outperforms the PED-ADPCM system. Selected audio examples coded by the presented error robust ADPCM coding schemes can be found for listening at [20].

4. CONCLUSIONS

The impact of the transmission errors on the base ADPCM using reconstructed signal driven (RSD) and prediction error driven (PED) lattice prediction error filters were discussed. An evaluation of latency free coding schemes were done to compare the perceptual audio quality for error-free and noisy channels.

The PED-ADCPM coding scheme guarantees high robustness against transmission errors up to BER of 10^{-4}. Beyond that, a fast transmission error decay in the decoded signal is ensured. On the other hand, good audio quality for error-free channels can only be achieved if a word length of at least 5 bit/samples is used. Consequently, the PED-ADPCM may be employed for real-time applications, where high error robustness is necessary and high channel bandwidth is available.

A transmission error robust RSD prediction filter in lattice structure was proposed. The proposed error robust RSD-ADPCM system similarly to the non-robust RSD-ADPCM achieves good audio quality for error-free channels. Additionally, the proposed method has an important robustness property against transmission errors. Although a complete decay of transmission error is much longer than in PED-ADPCM method, the proposed system guarantees up to BER of $5 \cdot 10^{-5}$ a slightly annoying impairment of the decoded signal. Therefore, an important step towards a latency free audio coding scheme with a good audio quality for a desired bit rate and error robustness is achieved.

5. REFERENCES

[1] Gayer Marc, Lutzky Manfred, Schuller Gerald, Krämer Ulrich, and Wabnik Stefan, "A guideline to audio codec delay," in *Audio Engineering Society Convention 116*, May 2004.

[2] 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.

[3] Markus Schnell, Markus Schmidt, Manuel Jander, Tobias Albert, Ralf Geiger, Vesa Ruoppila, Per Ekstrand, and Grill Bernhard, "MPEG-4 enhanced low delay AAC - a new standard for high quality communication," in *Audio Engineering Society Convention 125*, Oct. 2008.

[4] Martin Holters, Oliver Pabst, and Udo Zölzer, "ADPCM with adaptive pre- and post-filtering for delay-free audio coding," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, Apr. 2007, vol. 1, pp. I–269–I–272.

[5] 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.

[6] Martin Holters, Christian R. Helmrich, and Udo Zölzer, "Delay-free audio coding based on ADPCM and error feedback," in *Proc. ot the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Sep. 2008.

[7] D. Cohn and J. Melsa, "The residual encoder – an improved ADPCM system for speech digitization," Sep. 1975, vol. 23, pp. 935–941.

[8] International Telecommunication Union, "40, 32, 24, 13 kbit/s adaptive differential pulse code modulation ADPCM. ITU-T recommendation G.726," *General Aspects of Digital Transmission Systems, ITU-T*, 1990.

[9] Simon O. Haykin, *Adaptive Filter Theory*, Prentice Hall, 4 edition, Sept. 2001.

[10] P. Yatrou and P. Mermelstein, "Ensuring predictor tracking in ADPCM speech coders under noisy transmission conditions," *Selected Areas in Communications, IEEE Journal on*, vol. 6, no. 2, pp. 249–261, Feb 1988.

[11] D. Cohn and J. Melsa, "The relationship between an adaptive quantizer and a variance estimator (corresp.)," *Information Theory, IEEE Transactions on*, vol. 21, no. 6, pp. 669 – 671, Nov. 1975.

[12] J. Makhoul, "Stable and efficient lattice methods for linear prediction," Oct. 1977, vol. 25, pp. 423–428.

[13] 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.

[14] 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.

[15] Benjamin Friedlander, "Lattice filters for adaptive processing," *Proceedings of the IEEE*, vol. 70, no. 8, pp. 829–867, Aug. 1982.

[16] 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.

[17] Belgium European Broadcasting Union, Bruxelles, "Sound quality assessment material: Recordings for subjective tests – CD and user's handbook for the EBU-SQAM compact disc," Apr. 1988, Technical Centre of the EBU.

[18] International Telecommunication Union, "ITU recommendation ITU-R BS.1387-1, method for objective measurements of perceived audio quality (PEAQ)," Nov. 2001.

[19] P. Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," 2002, McGill University, Telecommunications & Signal Processing Laboratory.

[20] Audio examples, "Error robust delay-free lossy audio coding based on ADPCM," Jun. 2013, Available at http://ant.hsu-hh .de/dafx13/ADPCM/.