CASCADED PREDICTION IN ADPCM CODEC STRUCTURES

Marco Fink, Udo Zölzer

Department of Signal Processing and Communications,
Helmut-Schmidt University Hamburg
Hamburg, Germany
marco.fink@hsu-hh.de

ABSTRACT

The aim of this study is to demonstrate how ADPCM-based codec structures can be improved using cascaded prediction. The advantage of predictor cascades is to allow the adaption to several signal conditions, as it is done in block-based perceptual codecs like MP3, AAC, etc. In other words, additional predictors with a small order are supposed to enhance the prediction of non-stationary signals. The predictor cascade is complemented with a simple adaptive quantizer to yield a simple exemplary codec which is used to demonstrate the influence of the predictor cascade. Several cascade configurations are considered and optimized using a genetic algorithm. A measurement of the prediction gain and the ODG score utilizing the PEAQ algorithm applied to the SQAM dataset shall reveal the potential improvements.

1. INTRODUCTION

Many multimedia applications require high-quality streaming of audio content but only allow a restricted data rate for the transmission. This requirement led to the development of audio codecs which are successfully applied in manifold applications. However, there are some interactive musical applications, like wireless digital microphones or Networked Music Performances [1, 2], which additionally feature very strict latency requirements[3]. Popular audio codecs, like MP3, AAC, or HE-AAC, were designed to deliver lowest bit rates but feature algorithmic delays of up to hundreds of milliseconds. To decrease the delay contribution of audio codecs in interactive internet applications, the ultra-low delay codec [4, 5] and subsequently, OPUS [6, 7] were presented among others. Both codec approaches decrease the coding delay to a few milliseconds.

Codecs based on adaptive differential pulse code modulation (ADPCM), as analyzed in [8, 9, 10], cause a single sample delay which would be optimal for delay-sensitive applications but are ranked behind block-based methods in the quality-bitrate tradeoff. Robustness against transmission errors can be achieved by modifying the predictors as shown in [11]. To potentially improve the quality of ADPCM-like codecs, this study shall investigate how the application of cascaded predictors, as done for lossless coding in [12] and lossy coding in [4, 5], can yield increased prediction gains and therefore, higher perceptual quality. Note that the codec structures of [12, 4, 5] use cascaded prediction but are open loop implementations and the utilized ADPCM structure in this work is a closed loop approach.

Section 2 presents the structure of a typical ADPCM codec. The approach of a cascaded predictor is described in Sec. 3. The exemplary codec and its implementation is denoted in Sec. 4. The optimization of the cascade parameters is illustrated in Sec. 5 whereas the evaluation results are given in Sec. 6. Section 7 concludes this study.

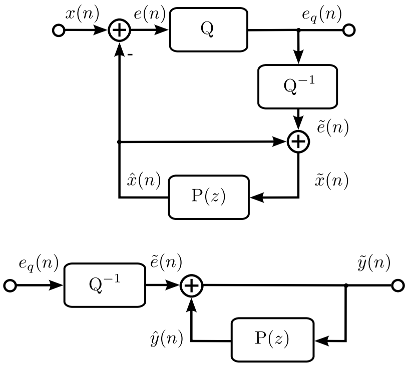

Figure 1: *Typical ADPCM encoder and decoder*

2. ADPCM

The main idea of an ADPCM codec is to solely quantize irredundant signal components. Therefore, a predictor $P(z)$ is applied to estimate a signal $\hat{x}(n)$ using old values of the original input signal $x(n)$. The resulting prediction error $e(n) = x(n) - \hat{x}(n)$ is then quantized, resulting in the quantized prediction error signal $e_q(n)$ which is actually transmitted. At the decoder side, the same predictor as in the encoder is applied to predict the signal $\hat{y}(n)$. The transmitted dequantized prediction error signal $\tilde{e}(n)$ is added to obtain the decoder output signal $\tilde{y}(n)$.

If predictor and quantizer are adaptive, the technique described above is called ADPCM. A block scheme describing ADPCM similar to [8, 9, 10] is depicted in Fig. 1. Since the predictor coefficients are not transmitted, it must be guaranteed that the predictor adaption in encoder and decoder is synchronous. This can be achieved by feeding them with the very same input data. Therefore, the predictor in the encoder is fed with a reconstructed input signal $\tilde{x}(n) = \hat{x}(n) + \tilde{e}(n)$, where $\tilde{e}(n)$ is the dequantized prediction error signal. In other words, the predictor adaption in the encoder is subject to quantization as well. Thus, guaranteeing that $\tilde{x}(n)$ and $\tilde{y}(n)$ are identical. Psychoacoustic knowledge can be involved by applying a set of pre- and post-filters in encoder and decoder to realize noise shaping as shown in [8, 9].

3. CASCADED PREDICTION

Concatenating several predictors and feeding them with the prediction error signal of the corresponding previous predictor is de-

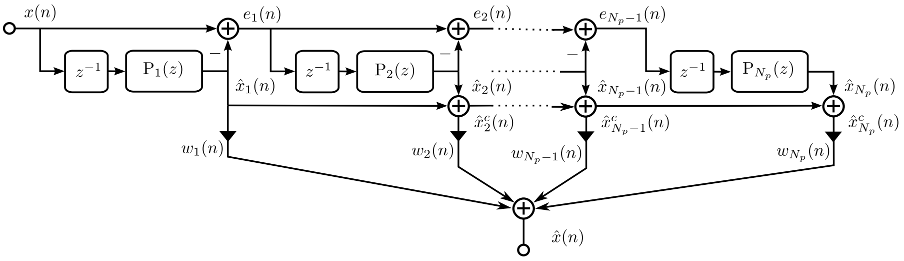

Figure 2: *Cascaded predictors*

noted a predictor cascade structure [13]. Figure 2 illustrates such a structure for N_p predictors. Apparently, the individual prediction signals $\hat{x}_p(n)$ are accumulated $\hat{x}_p^c(n) = \hat{x}_p(n) + \hat{x}_{p-1}^c(n)$ to produce an entire estimated signal per cascade stage.

As it can be seen in Fig. 2, the accumulated predictor outputs $\hat{x}_p^c(n)$ are multiplied with the weights $w_p(n)$ and summed to form the overall predictor output $\hat{x}(n)$. The weight $w_p(n)$ of predictor p is computed using the corresponding predictors prediction error signal

$$w_p(n) \propto e^{\left(-c(1-\mu)\sum\limits_{i=1}^{n-1}|e_p(n-i)|\cdot\mu^{i-1}\right)}, \qquad (1)$$

where $c = 2$ and $\mu = 0.9$ are tuning parameters [12].

This combination of a predictor cascade and *Predictive Minimum Description Length* (PMDL) weighting is called *Weighted Cascaded LMS* (WCLMS) and is used in lossless and lossy [4] codecs.

The advantage of this structure is that it allows to apply differently configured predictors. In other words, predictors of different order M_p and different adaption step sizes λ_p can be used in every cascade stage p, which can be optimized to certain signal characteristics. E.g. higher-order predictors, which adapt slowly, will nicely predict harmonic stationary sounds whereas fast-adapting low-order predictors are superior to follow non-stationary parts. Hence, the overall predictor is expected to adapt to a variety of signals and signal combinations.

An example is given in Fig. 3, where Fig. 3a) shows an excerpt of the trombone sample from the *Sound Quality Assessment Material* (SQAM) [14] dataset. A predictor of order $M_1 = 32$, a faster adapting second predictor of order $M_1 = 4$, and the corresponding cascade of order $M_{1/2} = [32, 4]$ are applied to this signal. The resulting recursively averaged prediction error energy

$$E(n) = (1-\alpha)\, e^2(n) + \alpha\, E(n-1) \qquad (2)$$

using $\alpha = 0.999$ is plotted in Fig. 3b). It can be seen that the higher-order predictor produces a clearly smaller prediction error signal than the lower-order predictor except for the first half second where the trombone signal mainly consists of noise until it produces a stable harmonic sound. In this section the smaller-order predictor is superior in terms of error energy. However, the cascade yields an overall result that almost combines the lower bounds of both configurations.

4. EXEMPLARY CODEC

The utilized simple codec for this study is basically structured as shown in Fig. 1. The predictor $P(z)$ is a cascade of lattice filters in this implementation. These lattice filters are adapted using the *Gradient Adaptive Lattice* (GAL) technique [15] and hence, the lattice predictor cascade is denoted *Weighted Cascaded Gradient Adaptive Lattice* (WCGAL) in the following. The implementation uses power-normalized adaption step sizes

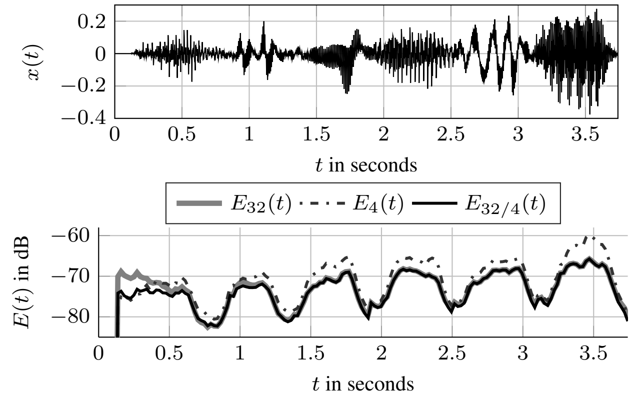

Figure 3: *Input signal and prediction error energy for predictors of order $M \in [32, 4]$ and their cascade. The test signal is a trombone sample from the SQAM dataset.*

$$\mu_m(n) = \frac{\lambda}{\sigma_m^2(n) + \sigma_{min}}, \qquad (3)$$

where m is the lattice stage index, n the sample index, σ_m^2 a recursive error power estimate, and σ_{min} a small offset to avoid a division by zero. λ is the base step size that is used as the main optimization parameter in the following.

The problem of optimally quantizing the prediction error signal $e(n)$ is not considered in this study and hence the quantizer Q is a simple fixed 3 bit quantizer with adaptive scaling. Normalizing the amplitude level is one way to achieve a nearly constant signal variance [16]. The normalization is accomplished by dividing the signal by an estimate of its envelope. The update itself is based on the denormalized quantized signal and hence, is synchronous in encoder and decoder. For more information about the envelope estimation and the calculation of the utilized quantizer levels, the interested reader is referred to [10].

5. OPTIMIZATION

Finding a meaningful combination of prediction order M_p and base step size λ_p for every predictor p of the cascade is a nontrivial task. The authors decided to apply a genetic algorithm to realize the global optimization of this problem.

Two metrics were considered to create potential cost functions. Initially, a cost function based on the prediction gain

$$G = 10\log_{10}\left(\frac{\sum\limits_{n=0}^{N-1} x(n)^2}{\sum\limits_{n=0}^{N-1} e(n)^2}\right), \tag{4}$$

defined as the logarithmic ratio of an input signal $x(n)$ of length N and the resulting (unquantized) prediction error signal $e(n)$, was used. But it turned out that a pure optimization of the prediction gain yields perceptually unpleasant results.

Therefore, a second cost function based on the so-called *Objective Difference Grade* (ODG) score was applied. The ODG score S is the outcome of the *Perceptual Evaluation of Audio Quality* (PEAQ) [17] method and describes the perceptual quality in terms of coding artifact audibility in a range from -4 (Very annoying) to 0 (Imperceptible). The actual utilized cost function

$$C(\boldsymbol{\chi}_{N_p}) = \sum_{k=1}^{70} S_k^4|\boldsymbol{\chi}_{N_p} \tag{5}$$

is the sum of the fourth power of all ODG scores $S_k^4|\boldsymbol{\chi}_p$ for all SQAM items k and a certain predictor cascade configuration $\boldsymbol{\chi}_{N_p}$. Raising S_k to the fourth power emphasizes bad results and hence this cost function tends to result in a globally enhanced ODG score instead of predominant excellent and some very poor results as explained in [18].

The optimization routine is repeated for several predictor cascade sizes N_p, a population size of 40, and $15 \cdot N_p$ generations. The range of valid values for the basic step size was restricted to $\lambda \in [1e^{-3}, 0.2]$. The trend of the cost function $C(\boldsymbol{\chi}_{N_p})$ for the predictor configurations over the generations of the genetic algorithm is illustrated in Fig. 4.

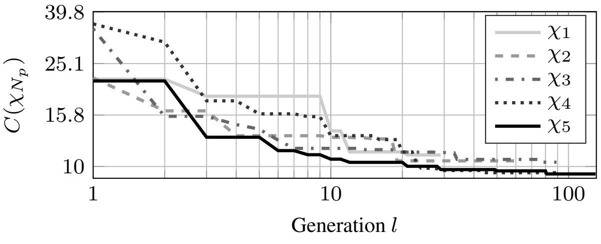

Figure 4: *Cost function $C(\boldsymbol{\chi}_{N_p})$ for cascade parameter vectors $\boldsymbol{\chi}_{N_p}$ over the generations l of the genetic algorithm*

For every generation l the best performing candidate of the population is shown. One can see how $C(\boldsymbol{\chi}_{N_p})$ is decreasing drastically over the generations. Apparently, the smallest cost function value can be achieved with the highest cascade size and vice-versa. Thereby, the expected gain through the application of a predictor cascade is proven. The cascade configurations and the associated results of the optimization process are denoted in Tab. 1. It denotes for the analyzed configurations c the optimized prediction order M_p, and the optimized adaption base step sizes λ_p.

Note that the genetic algorithm implementation from the MATLAB optimization toolbox with standard settings is used besides the mentioned parameters.

$\boldsymbol{\chi}_{N_p}$	= (Orders M_p	adaption base step size λ_p)
$\boldsymbol{\chi}_1$	(67,	0.0189)
$\boldsymbol{\chi}_2$	(50,4,	0.0020, 0.0423)
$\boldsymbol{\chi}_3$	(58,2,7	0.0016, 0.0098, 0.1036)
$\boldsymbol{\chi}_4$	(76,6,2,2	0.0013, 0.0119, 0.0848, 0.0549)
$\boldsymbol{\chi}_5$	(78,2,2,2,2	0.0010, 0.0199, 0.0779, 0.0052, 0.0024)

Table 1: *Optimized parameter vectors $\boldsymbol{\chi}_{N_p}$ of the prediction cascade for several cascade sizes N_p*

6. EVALUATION

To evaluate the WCGAL-based ADPCM codec structure, the same metrics as in Section 5 applied to the SQAM data set are utilized. The measurements are undertaken using the optimized values from Tab. 1.

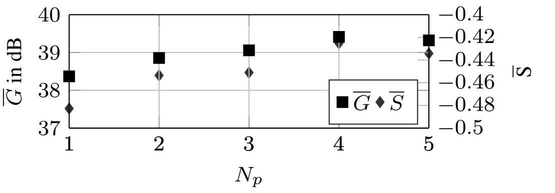

Figure 5: *Prediction gain $\overline{G}$ and ODG scores $\overline{S}$ averaged over all SQAM items for the settings of Tab. 1*

To get an impression of the overall performance, the averaged prediction gain $\overline{G}$ and ODG scores $\overline{S}$ for all configurations from Tab. 1 are illustrated in Fig. 5. Being contrary to the trend of the cost function, where its minimum value was found for the largest cascade size $N_p = 5$, the optimal cascade size in terms of mean ODG scores for the undertaken optimization is $N_p = 4$. In comparison to the single predictor configuration a gain of 0.06 for the mean ODG score and gain of about 1 dB for the mean prediction gain can be achieved.

Analyzing the individual SQAM items by plotting the prediction gain and the ODG score relative to the single predictor configuration $\boldsymbol{\chi}_1$, as done in Fig. 6, reveals the cause for the moderate gain. Despite the cost functions (see Eq. 5) intention of global perceptual enhancement, the individual gains are very signal dependent and occasionally (e.g. $k = [11, 32, 49, 53, 54]$) the predictor cascade even degrades the codecs performance. The mentioned negative examples are the double bass, the triangle, and 3 speech items and hence, a possible explanation could be the noise-like characteristic of those signals.

7. CONCLUSION

The application of cascaded predictors in ADPCM was analyzed in this work. In contrast to its previous application, the WCLMS concept was utilized for gradient-adaptive lattice filters (WCGAL) and in a closed loop codec structure. Order and adaption base step size of predictors of the cascade were optimized using a genetic algorithm by minimizing a cost function, depending on the perceptually motivated ODG score. A simple basic codec was implemented to evaluate this concept. The results for this non-ideal codec already indicate the benefit of cascaded prediction in an ADPCM codec.

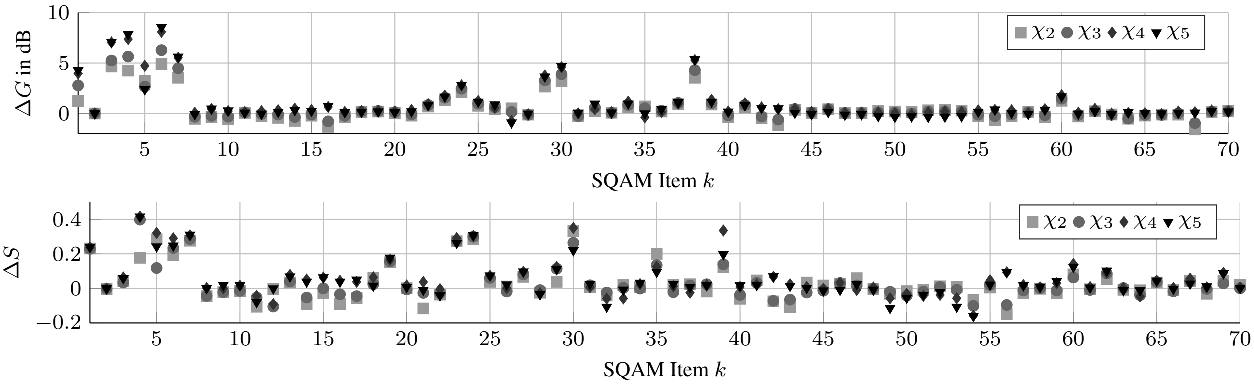

Figure 6: *Prediction gain a) and ODG score gain b) for every SQAM item utilizing the settings of Tab. 1. The results are relative to the ones for* $\boldsymbol{\chi}_1$.

A gain of about 0.06 for the mean ODG score and 1 dB for the mean prediction gain could be achieved without any change of the codecs bitrate. Unfortunately, the presented codec in conjunction with the utilized genetic optimization algorithm could not achieve a global optimization. In other words, the perceptual quality for some items of the used data base degraded. Hence, applying genetic algorithms might not be the optimal solution to find the best predictor cascade configuration.

The application of different optimization approaches, followed by a global optimization of all codec parameters potentially leads to an ADPCM codec offering a very good perceptual quality but featuring algorithmic delay of a single sample. Such a codec can be beneficial in many time-critical applications.

8. REFERENCES

[1] A. Carôt and C. Werner, "Network music performance-problems, approaches and perspectives," in *Proceedings of the Music in the Global Village*, Budapest, Hungary, 2007.

[2] A. Renaud, A. Carôt, and P. Rebelo, "Networked music performance: State of the art," in *Proceedings of the AES 30th International Conference*, Saariselkä, Finland, 2007.

[3] A. Carôt, C. Werner, and T. Fischinger, "Towards a Comprehensive Cognitive Analyisis of Delay-Influenced Rhytmical Interaction," in *Proceedings of the International Computer Music Conference (ICMC 2009)*, Montreal, Canada, 2009.

[4] J. Hirschfeld, J. Klier, U. Kraemer, G. Schuller, and S. Wabnik, "Ultra low delay audio coding with constant bit rate," in *AES Convention 117*, San Francisco, USA, Oct 2004.

[5] J. Hirschfeld, U. Kraemer, G. Schuller, and S. Wabnik, "Reduced bit rate ultra low delay audio coding," in *AES Convention 120*, Paris, France, May 2006.

[6] J.M. Valin, G. Maxwell, T. Terriberry, and K. Vos, "High-Quality, Low-Delay Music Coding in the Opus Codec," in *AES Convention 135*, New York, USA, 2013.

[7] J.M. Valin, T. Terriberry, C. Montgomery, and G. Maxwell, "A High-Quality Speech and Audio Codec With Less Than 10-ms Delay," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 1, Jan. 2010.

[8] M. Holters, O. Pabst, and U. Zölzer, "ADPCM with Adaptive Pre- and Post-Filtering for Delay-Free Audio Coding," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007)*, Honolulu, USA, April 2007.

[9] M. Holters and U. Zölzer, "Delay-free lossy audio coding using shelving pre- and post-filters," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, Las Vegas, USA, March 2008.

[10] M. Holters, C.R. Helmrich, and U. Zölzer, "Delay-Free Audio Coding Based on ADPCM and Error Feedback," in *Proc. of the 11th Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, 2008.

[11] G. Simkus, M. Holters, and U. Zölzer, "Error resilience enhancement for a robust adpcm audio coding scheme," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, May 2014.

[12] G. Schuller and A. Hanna, "Low delay audio compression using predictive coding," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2002)*, Orlando, USA, May 2002.

[13] P. Prandoni and M. Vetterli, "An FIR Cascade Structure for Adaptive Linear Prediction," *IEEE Transactions on Signal Processing*, vol. 46, September 1998.

[14] European Broadcast Union, "EBU Tech. 3253-E: Sound quality assessment material," April 1988.

[15] Lloyd J. Griffiths, "A continuously-adaptive filter implemented as a lattice structure," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '77)*, Hartford, USA, May 1977.

[16] D. Cohn and J. Melsa, "The relationship between an adaptive quantizer and a variance estimator (corresp.)," *IEEE Transactions on Information Theory*, vol. 21, Nov 1975.

[17] International Telecommunication Union, "ITU Recommendation ITU-R BS.1387: Method for objective measurements of perceived audio quality," November 2001.

[18] M. Holters and U. Zölzer, "Automatic parameter optimization for a perceptual audio codec," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, Taipei, Taiwan, April 2009.