ADPCM WITH ADAPTIVE PRE- AND POST-FILTERING FOR DELAY-FREE AUDIO CODING

Martin Holters, Oliver Pabst, Udo Zölzer

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

ABSTRACT

We propose an extension of ADPCM that includes adaptive pre- and post-filtering to achieve spectral shaping of the coding noise. The advantage of this coding scheme is that it allows a realization without algorithmic delay by making the filters backwards-adaptive. The measurements we present indicate that the addition of adaptive pre- and post-filtering to ADPCM results in a significant improvement in perceived audio quality. We therefore believe that the proposed system is a viable way to near-transparent lossy audio coding without algorithmic delay.

Index Terms— Audio coding, linear predictive coding, adaptive filters

1. INTRODUCTION

Over the past decades, lossy audio coding which exploits human auditory perception to achieve transparent quality at surprisingly low bit-rates has become ubiquitous. Most of the coding schemes are transform-based, making use of the energy compaction of time-to-frequency transforms and also employing psychoacoustic models in the frequency domain [1]. While these approaches have proved highly efficient in terms of bit-rate and achieved quality, the block-based transforms they employ inescapably cause high algorithmic delays which typically render them unusable for e.g. wireless microphones or in-ear monitors in live-performances. Coding schemes with low processing delays are typically aimed at speech coding e.g. for mobile phones and do not deliver satisfactory quality for other audio signals.

1.1. Ultra-low-delay coding

Only recently, audio coding systems with ultra low delay of 8 ms or less have been proposed [2, 3, 4]. The general structure of the systems is depicted in Fig. 1. The input signal is first fed through a filter which is adapted so that its transfer function resembles the inverse masking threshold of the input signal. The resulting signal, which is now normalized with respect to its masking threshold, is subsequently quantized and

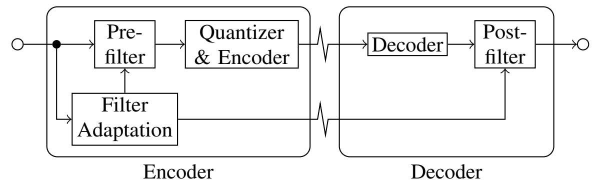


Figure 1: Overview of previous ultra-low-delay audio coding systems by Schuller *et al.*

encoded in a way that exploits redundancy to reduce quantization noise and/or required bit-rate. The quantization noise introduced in this step is approximately white. In the decoder, the signal is restored from the transmitted quantization indexes and filtered again to reconstruct the original spectrum. The filter coefficients required for this are transmitted as side information. As the post-filter of the decoder is the inverse to the pre-filter of the encoder, its magnitude resembles the masking threshold of the original signal and the formerly white quantization noise is shaped according to the masking threshold after the post-filter.

As it is not feasible to compute and transmit new coefficients for every sample, again a block-based approach is chosen. However, the blocks (and the introduced algorithmic delay) can be significantly smaller than for transform-based codecs. Thus, the ultra-low-delay codec is very suitable for use in packet-oriented networks, where a block-based transmission is unavoidable, anyway [5].

1.2. The proposed delay-free system

In the present paper, we propose a coding scheme where the pre-filter is backward-adapted from the signal after quantization, as shown in Fig. 2. As both encoder and decoder have access to this signal, it is no longer necessary to transmit the filter coefficients as side information and sample-wise processing without any algorithmic delay becomes possible. The pre-filtered signal is encoded using adaptive differential pulse code modulation (ADPCM) to take advantage of correlations still present in the signal after pre-filtering.

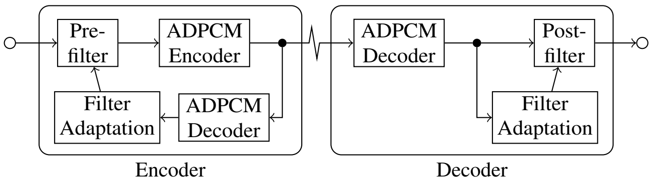


Figure 2: Overview of the proposed coding system.

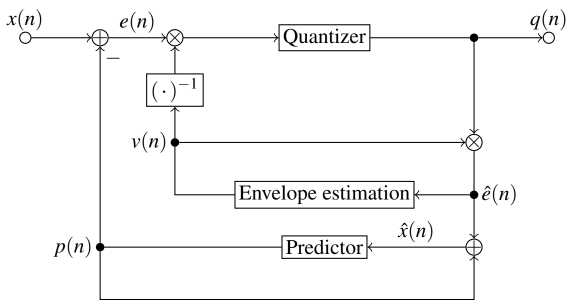


Figure 3: Structure of the ADPCM encoder.

2. ADPCM

The focus in the present paper is on the effect of the pre- and post-filtering, so we employ a simple and straight-forward implementation of sample-by-sample ADPCM. The general structure is shown in Fig. 3. First, a prediction $p(n)$ is subtracted from the input signal $x(n)$ of the ADPCM encoder. The resulting prediction error $e(n) = x(n) - p(n)$ is then multiplied by the reciprocal of an estimated envelope $v(n)$ to yield a signal with a nearly constant amplitude envelope, which is subsequently quantized down to yield the desired bit-rate. The resulting low bit-rate signal $q(n)$ is transmitted to the decoder, and also used for the prediction and envelope estimation.

To estimate the amplitude envelope of the prediction error $e(n)$, the absolute value of the reconstructed prediction error $\hat{e}(n) = v(n) \cdot q(n)$ is low-pass filtered as shown in Fig. 4. The parameter b controls the attack time of the envelope estimator, where lower values of b result in shorter attack times. The filter is designed such that its output is four times larger than the mean of its input, such that for typical distributions of $e(n)$, a reasonably low number of samples of $e(n)/v(n)$ will have an absolute value exceeding 1. To avoid application of arbitrarily large gains to the prediction error, the envelope is lower-bounded by enforcing $v(n) \geq v_{min}$.

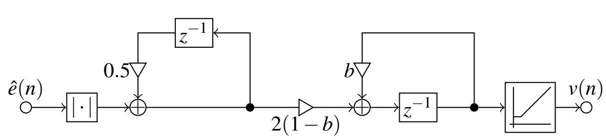


Figure 4: The envelope estimator used in the ADPCM encoder.

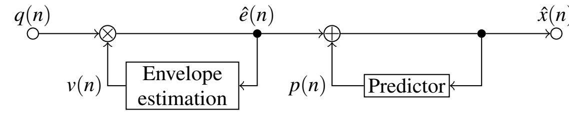


Figure 5: Structure of the ADPCM decoder.

The prediction is obtained from the reconstructed signal $\hat{x}(n) = \hat{e}(n) + p(n)$ by applying a standard least mean squares (LMS) predictor, using an Nth order FIR prediction filter

$$p(n) = \sum_{m=1}^{N} h_m(n) \cdot \hat{x}(n-m) \tag{1}$$

and an update scheme

$$h_m(n+1) = h_m(n) + \mu \cdot \hat{e}(n) \cdot \hat{x}(n-m) \tag{2}$$

where the filter coefficients h_m are updated only from data already present in the filter [6]. Note that neither envelope estimator nor predictor contain direct paths, so that no delay-free loops occur in the ADPCM encoder.

The quantizer is implemented by rounding to the nearest multiple of 2^{w-1} in the range $[-1, 1]$, where w is the desired number of bits per sample. Clearly, a non-uniform quantization would yield better results, but for reasons of brevity, we have chosen a uniform quantizer.

Reconstruction of the signal in the decoder works in the same way as in the encoder, where the reconstructed signal was used to feed the predictor, as shown in Fig. 5. By construction, both encoder and decoder always compute the same values for $v(n)$ and $p(n)$.

3. PRE- AND POST-FILTERING

Ideally, the post-filter's magnitude response should resemble the masking threshold of the audio signal, and the pre-filter its inverse. However, with the closed-loop structure of the encoder, it is important to have a filter adaptation without significant delays, to adapt quickly to changed signal characteristics and to avoid instabilities or oscillations.

We have therefore chosen not to compute the masking threshold explicitly. Instead, we make use of the heuristic that the masking is basically a smoothened version of the signals spectrum and that the desired pre-filter whitens the input signal to a certain extent. This allows the application of linear prediction methods in this step as well.

In particular, we choose the pre-filter to be an all-pole filter

$$x(n) = u(n) - \sum_{m=1}^{M} w_m(n) \cdot x(n-m), \tag{3}$$

where $u(n)$ denotes the filter input and $x(n)$ its output. The corresponding post-filter in the decoder then of course is given

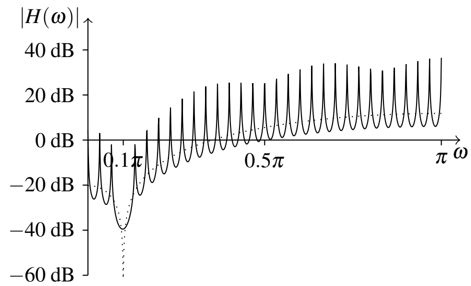


Figure 6: Magnitude responses of prediction error filters when used to predict a single sinusoid at $\omega = 0.1\pi$ for an all-pole filter (solid line) and a zero-only filter (dotted line).

by the FIR filter

$$\hat{u}(n) = \hat{x}(n) + \sum_{m=1}^{M} w_m(n) \cdot \hat{x}(n-m), \tag{4}$$

where $\hat{x}(n)$ is the result of coding and decoding $x(n)$ with ADPCM and $\hat{u}(n)$ denotes the reconstructed signal of the proposed system.

Usage of an all-pole filter as pre-filter has the advantage that even for very tonal signals, the magnitude response of the pre-filter will have a limited attenuation at the respective frequencies. In contrast to this, a filter with at least one complex conjugate pair of zeros would yield a magnitude response with a very sharp notch when used to predict a single sinusoid, as shown in Fig. 6. Although the ripple caused by the individual poles is objectionable, the overall result when the inverse filter shapes the quantization noise is still better than the strong noise peak the inverse of the filter with zeros would produce at the frequency of the sinusoid.

The filter is adapted with a variant of the leaky signed LMS algorithm [7]. In the usual form, the filter coefficients $w_m(n)$, $m = 1, \ldots, M$ would be updated by

$$w_m(n+1) = \alpha w_m(n) + \beta \operatorname{sign}\big(x(n) \cdot x(n-m)\big) \tag{5}$$

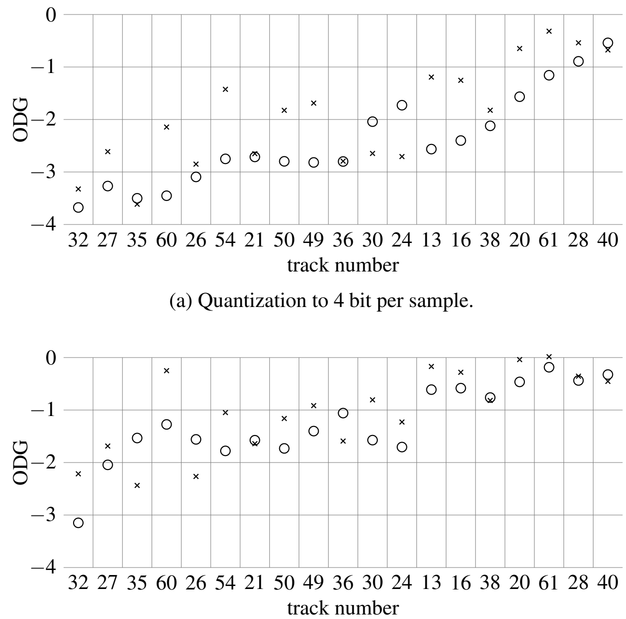
with step-size β and leakage parameter α. To have exactly the same filter coefficients as for the post-filter in the decoder, however, we replace all occurrences of $x(n)$ in equation (5) with $\hat{x}(n)$, the output after encoding and decoding with ADPCM. Thus, the filter coefficients are updated according to

$$w_m(n+1) = \alpha w_m(n) + \beta \operatorname{sign}\big(\hat{x}(n) \cdot \hat{x}(n-m)\big) \tag{6}$$

in both encoder and decoder.

4. RESULTS

To evaluate the effects of the pre- and post-filtering, we have coded a number of test signals with both plain ADPCM and the same ADPCM with additional pre- and post-filtering. To compare the results, we determined the objective difference



(a) Quantization to 4 bit per sample.

(b) Quantization to 6 bit per sample.

Figure 7: The resulting ODGs after encoding and decoding with plain ADPCM (circles) and ADPCM with adaptive pre-/post-filtering (crosses).

grade (ODG) according to [8]. While objective evaluation cannot completely replace listening tests, the results may well be used to identify trends when comparing similar codecs. The test signals were selected tracks from the SQAM CD [9].

The filter order of the ADPCM codec was chosen as $N = 30$, the step-size for the adaptation of the prediction filter was set to $\mu = 0.1$. For the low-pass of the envelope estimator, $b = 0.9789$ was used, and the result was lower-bounded to $v_{min} = 2^{-10}$. The resulting ODGs when coding and decoding with this simple ADPCM are denoted with circles in Fig. 7.

As expected, the audio quality of the plain ADPCM is rather poor. Especially when coding to 4 bit per sample, for the majority of the test signals the measured ODG is worse than -2, and only two signals are reconstructed with an ODG better than -1. Naturally, audio quality improves significantly when the bit-rate is increased to 6 bit per sample. Now except for only two test signals, the ODGs exceed -2, and for about one third of the signals, it is better than -1.

The pre- and post-filters we added to the system were of order $M = 150$ and the adaptation parameters were set to $\alpha = 0.999$ and $\beta = 0.001$, respectively. The measured audio quality, depicted by crosses in Fig. 7, shows a clear improvement for most signals. For 4 bit per sample, the number of signals with an ODG worse than -2 is reduced form 14 to 9, while the number of signals with an ODG better than -1 is doubled from 2 to 4. The average improvement of the ODG is 0.48.

For a quantization to 6 bit per sample, the improvement is less pronounced, but still present. The number of signals for which the ODG exceeds -1 is increased from 7 to 10, while unfortunately, the number of signals leading to an ODG below -2 is increased by 1 to 3. On average, the ODG is improved by 0.23. It is worth noting that the three signals where adding the pre- and post-filter leads to a significant decrease of audio quality are claves, glockenspiel and xylophone — signals that exhibit tonal signals with sharp onsets. For these, the additional filtering leads to an impairment of the transients which is not justified by the improvement for the stationary parts.

5. CONCLUSIONS

We have presented adaptive pre- and post-filtering as a simple way to improve the quality of ADPCM, aiming at a delay-free audio compression scheme. From the results presented, it is obvious that despite significant improvements in audio quality for the simple ADPCM codec employed, the achieved quality is still unsatisfactory.

However, it has to be considered that improvements of the ADPCM itself, like more sophisticated adaptation of the prediction filter and a non-uniform quantizer, will likely be complementary to the effects of the pre- and post-filtering. In particular, modifications of the ADPCM part may aim at reduction of the overall noise level, especially for stationary parts, or faster adaptation for non-stationarities, where often a trade-off between these two goals is required. The pre- and post-filtering, on the other hand, aims at spectrally shaping the coding noise, even allowing an increase of the noise power while at same time improving perceived quality.

Thus, the use of adaptive pre- and post-filtering may allow the ADPCM codec to be tuned more towards fast adaptation, sacrificing some of the stationary case performance. This kind of joint optimization was not performed for the results presented in this paper, where the focus is on the effects of the pre- and post-filtering itself. We therefore believe that by employing more advanced ADPCM and jointly optimizing all parameters of the proposed system, near-transparent lossy audio coding without algorithmic delay is possible.

6. REFERENCES

[1] T. Painter and A. Spanias, “Perceptual coding of digital audio,” *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000.

[2] Bernd Edler and Gerald Schuller, “Audio coding using a psychoacoustic pre- and post-filter,” in *Proc. ICASSP 2000*, Istanbul, June 2000, IEEE, vol. 2, pp. 881–884.

[3] Gerald Schuller and Aki Härmä, “Low delay audio compression using predictive coding,” in *Proc. ICASSP 2002*, Orlando, May 2002, IEEE, vol. 2, pp. 1853–1856.

[4] Stefan Wabnik, Gerald Schuller, Jens Hirschfeld, and Ulrich Krämer, “Reduced bit rate ultra low delay audio coding,” in *Proceedings of the 120th AES Convention*, Paris, May 2006, AES.

[5] Alexander Carôt, Ulrich Krämer, and Gerald Schuller, “Network music performance (NMP) in narrow band networks,” in *Proceedings of the 120th AES Convention*, Paris, May 2006, AES.

[6] Simon Haykin, *Adaptive Filter Theory*, Prentice Hall, Englewood Cliffs, third edition, 1995.

[7] David B. Thomas, “Signal predictive audio transmission system,” 2003, United States Pregrant US 2003/0191634.

[8] ITU Recommendation ITU-R BS.1387-1, “Method for objective measurements of perceived audio quality,” 2001.

[9] EBU Tech. 3253-E, “Sound quality assessment material. Recordings for subjective tests,” Apr. 1998.