

DELAY-FREE LOSSY AUDIO CODING USING SHELving PRE- AND POST-FILTERS

Martin Holters, Udo Zölzer

Helmut Schmidt University, Hamburg, Germany

ABSTRACT

A delay-free audio coding scheme based on ADPCM with adaptive pre- and post-filtering is presented. The pre-/post-filters are realized as a cascade of shelving filters, designed to match the characteristics of human perception. The pre- and post-filters are adapted by dynamic compression of the respective sub-bands. The adaption is backward-adaptive, i.e. is fed by the reconstructed signal, which eliminates the need to transmit the filter coefficients and allows delay-free operation. This pre- and post-filtering significantly improves the audio quality compared to a plain ADPCM codec, as underlined by objective measurements. Since the base ADPCM used is also delay-free, the resulting coding system works without any algorithmic delay.

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

1. INTRODUCTION

Bit-rate reduction of audio-data which exploits human auditory perception to dramatically reduce the amount of data to be transferred or stored without incurring perceptible distortion has become ubiquitous. Most of the coding schemes in use, including the well-known MP3, are transform-based and operate in the frequency-domain [1]. Without doubt, this has proved highly effective in terms of bit-rate reduction and audio quality. However, the block-based processing inherent of the transform-based codecs renders them unusable for applications where processing latency is a concern, like wireless microphones or in-ear monitors in live performances.

On the other hand, coding schemes with a special focus on low latency have often only targeted speech signals, being primarily developed for mobile phones, and perform worse than desirable for general audio signals. Only recently, approaches for low-delay audio coding have been proposed. In this paper, we present a codec without any algorithmic delay, aiming especially at live-performance applications.

2. PRIOR LOW-DELAY CODING SCHEMES

In [2], a coding system with a delay of 10 ms or less was proposed. Its general outline is depicted in Figure 1. The input signal is first filtered with a pre-filter which is adapted such that

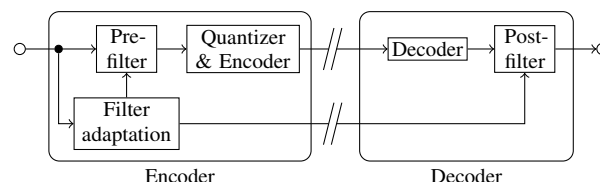


Fig. 1. Outline of the Fraunhofer ultra-low-delay audio codec.

it approximates the inverse hearing threshold in the presence of the original signal. The following quantization and encoding exploits any redundancy in the pre-filtered signal to reduce the bit-rate with minimum distortion. The error introduced in the process is approximately white. Finally, a post-filter which is inverse to the pre-filter restores the original magnitude spectrum, and shapes the quantization error according to the hearing threshold, reducing its audibility.

To allow the post-filter to be the exact inverse of the pre-filter, the filter coefficients are transmitted alongside the quantized audio data. The resulting increase in bit-rate is limited by cutting the audio stream into frames and transmitting the filter coefficients only once per frame. These frames are the main reason of algorithmic delays in the proposed coding schemes, but as they may be significantly smaller than in transform-based codecs, the delays are reduced substantially.

In [3] an alternative approach, LowPAC, was proposed where the signal is decomposed into eight sub-bands which are individually encoded using adaptive differential pulse code modulation (ADPCM). While the quantization noise in each sub-band is approximately white, its power is different across bands, depending on the signal power in each band, leading to the desired spectral shaping of the quantization noise. The main source of algorithmic delay in LowPAC are the analysis and synthesis filter banks, which cause about 3 ms total delay. This improvement in latency compared to the Fraunhofer ultra-low-delay codec comes at the cost of slightly worse audio quality, especially for very percussive signals.

3. THE PROPOSED SYSTEM

The audio coding scheme proposed in this paper follows the concept of [4], as depicted in Figure 2. Like the Fraunhofer ultra-low-delay codec, it consists of pre- and post-filters and

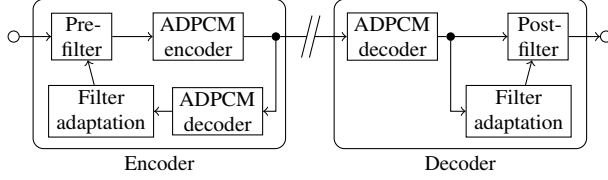


Fig. 2. Overview of the proposed coding system.

a broadband base codec, ADPCM in this case. The filters, however, are adapted from the reconstructed signal to avoid the need of coefficient transmission. As the signal does not need to be split up in frames, and the base ADPCM codec is latency-free, the coding system does not have any algorithmic delay.

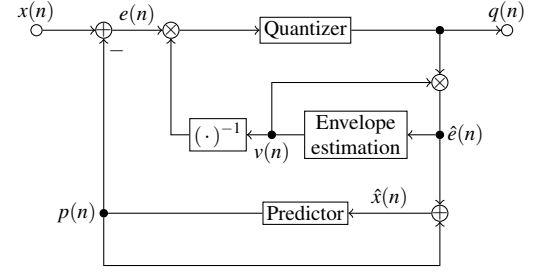
3.1. The base ADPCM codec

A fairly standard ADPCM encoder and decoder similar to the ones from [3] and [4] has been employed in this work, see Figure 3. From the input $x(n)$ to the ADPCM encoder, a prediction $p(n)$ is subtracted. The resulting prediction error $e(n) = x(n) - p(n)$ is then divided by an estimate $v(n)$ of its expected absolute value. This normalized prediction error is finally quantized to in our case 4 bit using a uniform quantizer.

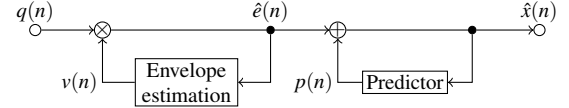
The value $q(n)$ associated with the 4 bit quantization index is then multiplied with $v(n)$ in both decoder and encoder to obtain the reconstructed prediction error $\hat{e}(n) = v(n) \cdot q(n)$. From this $\hat{e}(n)$, the level-estimate $v(n+1)$ to be used for the next sample is determined. By further adding the predicted value back on the prediction error, we obtain the reconstructed signal $\hat{x}(n) = \hat{e}(n) + p(n)$, again both in decoder and encoder. The predictor uses $\hat{x}(n)$ to calculate the prediction $p(n+1)$ for the next input sample. As the same values of $\hat{e}(n)$ and $\hat{x}(n)$ are used in encoder and decoder to update $v(n)$ and $p(n)$, respectively, it is not necessary to transmit any side-information, provided that the same initial states are used. The overall error introduced is approximately white, as can easily be verified.

The level or envelope estimation is performed by low-pass filtering $\hat{e}^2(n)$ and taking the square root of the result. This results in a signal at the quantizer's input which is normalized with respect to its power, so that limits of the quantizer can be chosen accordingly. To prevent overloading the quantizer when the prediction error power suddenly increases, the low-pass switches between two cut-off frequencies: when $\hat{e}^2(n) > v^2(n)$, a higher cut-off frequency is used to facilitate fast adaption, while otherwise, a lower cut-off frequency is used to have a smoother $v(n)$ for stationary signals.

The predictor is realized as an FIR lattice filter, the coefficients being adapted using the gradient adaptive lattice (GAL) algorithm [5]. The lattice filter structure has the important advantage that it can easily be ensured to be minimum-phase, a prerequisite for the stability of the feed-back structure employed. The GAL algorithm is very attractive because of its



(a) Encoder.



(b) Decoder.

Fig. 3. Structure of the ADPCM codec.

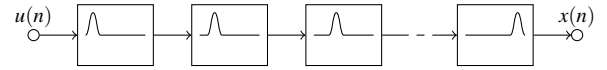


Fig. 4. The proposed pre-filter as a cascade of band-shelving filters.

low computational complexity.

This baseline ADPCM certainly leaves room for optimizations. Especially, use of a non-uniform quantizer promises significant gains in audio quality. Nevertheless, it will suffice to show that additional pre- and post-filtering are beneficial.

3.2. The pre-filter

The pre-filter employed in this new approach consists of a cascade of several band-shelving filters, see Figure 4. Each individual shelving filter applies a given gain in certain band. The bands are selected to match the 25 critical bands defined by Zwicker [6], with the lower limit frequency of the lowest band set to 10 Hz and the highest band reaching up to 20 kHz.

The band-shelving filters are an extension of peak filters to higher filter orders [7, 8]. They are used to realize the magnitude response

$$|H_g(e^{j\Omega})|^2 = \frac{(\cos \Omega_c - \cos \Omega)^M + (K \sin \Omega)^M g}{(\cos \Omega_c - \cos \Omega)^M + (K \sin \Omega)^M g^{-1}}, \quad (1)$$

where M denotes the necessarily even filter order and g the desired gain. The frequency band the shelving filter works in is specified by the normalized center frequency

$$\Omega_c = 2 \arctan \left(\sqrt{\tan(\pi f_u / f_s) \cdot \tan(\pi f_l / f_s)} \right) \quad (2)$$

and a coefficient

$$K = \tan(\pi \cdot (f_u - f_l) / f_s) \quad (3)$$

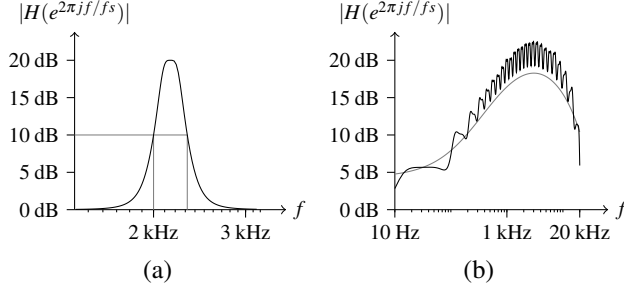


Fig. 5. Magnitude responses of fourth order shelving filters, (a) single filter for $f_l = 2000$ Hz, $f_u = 2320$ Hz and $g = 10$, (b) cascade of filters to approximate given magnitude response (gray).

depending on the bandwidth, where f_l and f_u denote lower and upper cut-off frequency, respectively, and f_s is the sampling frequency. With this selection of parameters, the magnitude response reaches the gain g at the center frequency and \sqrt{g} (half the gain in dB) at the band edges. An example of the resulting magnitude response when a gain $g = 10$ (20 dB) is applied in the 14th band using a fourth order band-shelving filter is shown in Figure 5a.

By cascading these band-shelving filters and choosing the gains appropriately, arbitrary magnitude responses can be approximated, as exemplified in Figure 5b. The frequency resolution is quite limited, but matches human auditory perception by construction, as the filter bands correspond to critical bands of perception. Choosing the correct gains is simplified by using higher-order filters compared to the traditional second-order peak-filters, as the influence of one filter outside its band is drastically reduced, so that the gains can be determined independently.

3.3. The post-filter

As the band-shelving filters are minimum-phase, they yield a stable inverse which can be used in the post-filter. In fact, for these filters $H_g^{-1}(z) = H_{1/g}(z)$, promising a very simple inversion by using the same filter with inverse gain. However, as the filters will be adapted by changing the gains g , things become a bit more complicated.

It is helpful to consider the state-space representation of the pre-filter

$$x(n) = \mathbf{c}^T(n)\mathbf{s}(n) + d(n)u(n) \quad (4)$$

$$\mathbf{s}(n+1) = \mathbf{A}(n)\mathbf{s}(n) + \mathbf{b}(n)u(n) \quad (5)$$

with the state vector $\mathbf{s}(n)$. Assuming $\mathbf{s}(n)$ is known to the inverse filter, of course (4) can simply be solved for $u(n)$ to give

$$u(n) = \frac{x(n) - \mathbf{c}^T(n)\mathbf{s}(n)}{d(n)}. \quad (6)$$

With this recovered $u(n)$, it is then possible to update the copy of the state-space vector in the inverse filter according to (5). Thus, the filter defined by (6) and (5) is the exact inverse of the filter given by (4) and (5) even when the coefficients change over time.

To apply this principle in practice, it is not necessary to explicitly obtain the state-space representation. It suffices to determine $d(n)$ and $\mathbf{c}^T(n)\mathbf{s}(n)$ to be able to evaluate (6) and then use the same state update mechanism as in the pre-filter. A further simplification is possible as the individual sections of the filter cascade can be inverted one by one, proceeding in reverse order. By following this approach, the computational complexity of pre- and post-filter are about the same.

3.4. Adaption of pre- and post-filter

Optimally, the magnitude response of the post-filter should resemble the hearing threshold under presence of the original signal, to achieve best masking of the quantization noise. In this paper however, we have chosen not to employ a full-featured psycho-acoustic model, as we have to recompute the filter coefficients with every sample. Instead, a dynamic compression is performed in each band by the encoder, and the corresponding expansion in the decoder produces the desired spectral shaping. This has been inspired by the promising results of spectral amplitude warping [9].

In particular, let $P_u(i)$ and $P_x(i)$ denote the power of the signal before and the pre-filter, respectively, falling into the i th band ($i = 1, \dots, 25$). Then we want

$$P_x(i) = P_u^\alpha(i), \quad (7)$$

where $0 < \alpha < 1$ specifies the amount of dynamic compression, with lower values resulting in higher compression. If we idealize the shelving filters to have infinitely steep edges, we have

$$P_x(i) = g^2(i) \cdot P_u(i) \quad (8)$$

with gain $g(i)$ in the i th band, and therefore have to choose

$$g(i) = P_u^{\frac{\alpha-1}{2}}(i). \quad (9)$$

The $P_u(i)$ are determined by decomposing the signal into the respective sub-bands and applying a level estimator like described in section 3.1. The analysis filters are Butterworth band-passes matching the respective shelving filters in order and cut-off frequencies. To avoid transmission of the $g(i)$, the adaption again is based on the reconstructed signal $\hat{u}(n)$ instead of $u(n)$.

4. EVALUATION

To evaluate the impact of the pre-/post-filter, objective audio quality measurements were performed for the plain ADPCM codec and the ADPCM codec with pre-/post-filtering. The

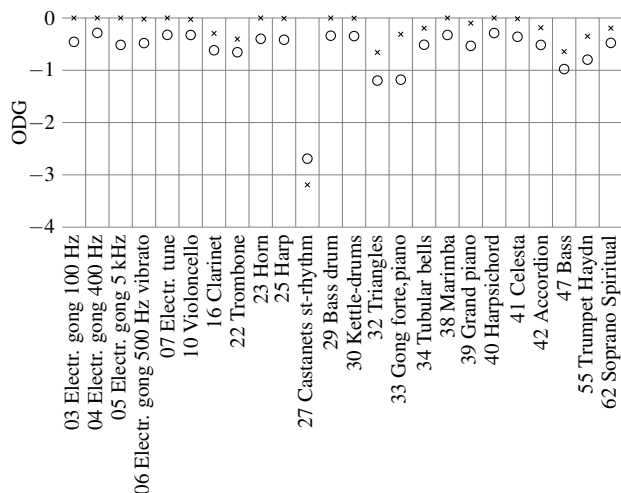


Fig. 6. Comparison of plain ADPCM (circles) and ADPCM with pre-/post-filtering (crosses) in terms of the ODG.

audio quality was determined by the PEAQ method [10, 11], resulting in an objective difference grade (ODG) ranging from -4 (very annoying) to 0 (imperceptible). The test items were mono down-mixes of the EBU sound quality assessment material [12] with a sample-rate of 44.1 kHz, giving a coded bit-rate of 176.4 kbit/s.

In Figure 6 the ODGs produced by the plain ADPCM codec along with those after adding the pre-/post-filter are given for those items where the difference exceeds 0.25. The shaping filters used are of order $M = 4$, the compression exponent was chosen as $\alpha = 0.86$.

For most of the test items, adding the pre-/post-filter results in an improvement in audio quality which was confirmed in informal listening tests. The notable exception is item 27 where the pre-/post-filter actually degrades quality. Item 27 is very percussive and hence very problematic for the backwards-adaptive scheme proposed, where the extra adaption of the pre-/post-filtering just adds to the problem.

Nevertheless, the mean ODG over all test items is improved from -0.579 to -0.348 , representing a clear improvement in quality. Considering that the ADPCM codec is somewhat constrained by using a uniform quantizer and thus leaves room for optimization, the achieved audio quality already is very respectable.

5. CONCLUSIONS

In [4], the authors proposed a coding scheme for delay-free coding using a backwards-adaptive pre-/post-filtering scheme. The proposition was substantiated by results showing that even an ad-hoc approach for the pre-/post-filter could improve the audio quality of an otherwise poor ADPCM codec.

With this paper, the authors now show that using an im-

proved ADPCM codec and a systematically designed pre- and post-filter, audio data can be coded to 4 bit per sample at acceptable quality levels without any algorithmic delay. With further improvements to the base ADPCM codec, like using a non-uniform quantizer, it may be expected that the proposed scheme is a viable approach for delay-free, near-transparent lossy audio coding.

6. REFERENCES

- [1] Ted Painter and Andreas Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000.
- [2] Gerald Schuller, Bin Yu, Dawei Huang, and Bernd Edler, "Perceptual audio coding using adaptive pre- and post-filters and lossless compression," *IEEE Trans. Speech Audio Process.*, vol. 10, no. 6, pp. 379–390, Sept. 2002.
- [3] Florian Keiler, "Real-time subband-ADPCM low-delay audio coding approach," in *Proc. 120th AES Convention*, Paris, May 2006, AES, Paper 6748.
- [4] Martin Holters, Oliver Pabst, and Udo Zölzer, "ADPCM with adaptive pre- and post-filtering for delay-free audio coding," in *Proc. ICASSP '07*, Honolulu, Apr. 2007, IEEE.
- [5] Lloyd J. Griffiths, "A continuously-adaptive filter implemented as a lattice structure," in *Proc. ICASSP '77*, Boulder, May 1977, IEEE.
- [6] Eberhard Zwicker and Hugo Fastl, *Psychoacoustics*, Springer, Berlin, second edition, 1999.
- [7] Sophocles J. Orfanidis, "High-order digital parametric equalizer design," *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1026–1046, Nov. 2005.
- [8] Martin Holters and Udo Zölzer, "Parametric recursive higher-order shelving filters," in *Proc. 120th AES Convention*, Paris, May 2006, AES, Paper 6722.
- [9] Roch Lefebvre and Claude Laflamme, "Spectral amplitude warping (SAW) for noise spectrum shaping in audio coding," in *Proc. ICASSP '97*, Munich, Apr. 1997, IEEE, vol. 1, pp. 335–338.
- [10] ITU Recommendation ITU-R BS.1387-1, "Method for objective measurements of perceived audio quality," 2001.
- [11] Peter Kabal, "An examination and interpretation of ITU-R BS.1387: Perceptual evaluation of audio quality," Tech. Rep., Department of Electrical & Computer Engineering, McGill University, Montreal, 2003.
- [12] EBU Tech. 3253-E, "Sound quality assessment material. Recordings for subjective tests," Apr. 1998.