GLOBALLY OPTIMIZED DYNAMIC BIT-ALLOCATION STRATEGY FOR SUBBAND ADPCM-BASED LOW DELAY AUDIO CODING

Stephan Preihs and Jörn Ostermann

Leibniz Universität Hannover, Institut für Informationsverarbeitung
Appelstr. 9A, 30167 Hannover, Germany, {preihs, ostermann}@tnt.uni-hannover.de

ABSTRACT

This paper presents an extension and global optimization of a bit-allocation strategy for use in a low delay subband ADPCM-based audio coding scheme. The framework employed for the assignment of bits to subband quantizers is based on a subband level estimation of the signals used for the prediction error normalization. It is modified to reduce switching artifacts and extended to contain a mapping to predefined allocation presets.

Since our bit-allocation scheme as well as the subband coding involve several partially interacting parameters that are hard to adjust manually, a framework for their global optimization is presented.

Experiments and their results show that in our coding scheme a significant improvement of audio quality without large signaling overhead can be achieved by the use of the modified bit-allocation. Furthermore, the global optimization allows for an additional gain in audio quality compared to manually adjusted parameters.

Index Terms— Low delay audio coding, dynamic bit-allocation, global optimization techniques, subband ADPCM

1. INTRODUCTION

Low delay audio coding is required when real time audio transmission systems have to deal with limited data rates. When it comes to wireless audio transmission in a live scenario with a high number of simultaneous channels, the constraints on delay and bandwidth are different compared to the usual ones. In addition, some applications take place at the beginning of the "audio production chain" and therefore require a near transparent audio quality which leads to a demand for low delay source coding algorithms with delays less than $1\,\mathrm{ms}$ and a very high audio quality.

This makes the well known and established low delay codecs like AAC-ELD [1] and FhG ULD [2] or even the CELT (Opus) codec [3] unsuitable in these scenarios. One possible approach for audio coding with very low delay and near transparent audio quality is the one we presented in [4]. There, we combined a numerically optimized filter bank designed with a variant of the framework presented in [5] with an error robust version of the ADPCM presented in [6].

In the encoder an analysis filter bank splits the input signal into five critically downsampled subband signals which are processed by an ADPCM-based time-domain coding. In the decoder the subband signals are reconstructed and a synthesis filter bank generates the reconstructed output signal. Since the coding of subband signals is inherently delay free, the delay of the codec is mainly dependent on the group delay of the filter bank and the, in a real-world system, necessary sequential transmission of quantization indices.

While in [4] we were able to show that our coding scheme allows for a near transparent audio quality for the majority of test signals, for some critical signals there is still a need for improvement. Therefore we did extensive research on how to improve our error robust ADPCM-based coding of subband signals. In addition we found that a static assignment of bits to the subband quantizers prevents from a sufficient audio quality for some signals. Following this, we have been looking for a method for dynamically assigning bits to the subband quantizers which does not lead to an additional delay of the codec or a significant overhead and has a rather low computational cost.

In [7] we found a bit-allocation strategy that is both simple and effective and, according to the authors, is practically similar to classic assignment criteria like the ones contained in [8] and [9]. Using this approach in our coding scheme we found that it results in unnecessary overhead and leads to a suboptimal adaption to signal changes which can cause audible switching artifacts. Therefore we propose appropriate modifications in the following.

2. DYNAMIC BIT-ALLOCATION STRATEGY

Figure 1 shows the basic structure of our modified coding scheme. The flow graph is extended to contain an allocation and mapping block (ALLOC&MAP) which has the prediction error normalization signals of the adaptive prediction and quantization blocks ($AdPQ_i$) as input and the mapped bit-allocation as output. In addition the subband scheme now contains a multiplexing and demultiplexing unit (MUX, DeMUX) for the generation and transmission of bit-allocation signaling information in a frame-based structure.

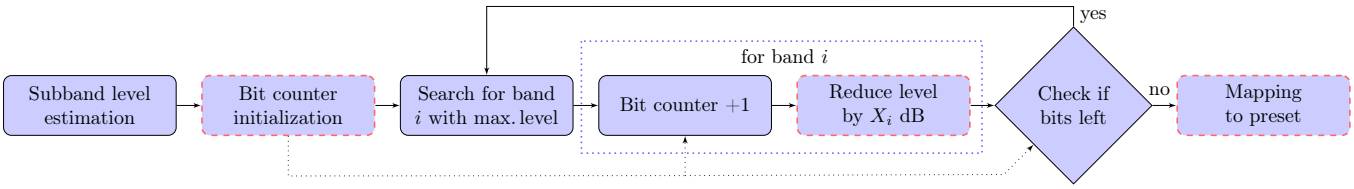

Fig. 2. Flow graph of the extended dynamic bit-allocation strategy (modified blocks dashed red).

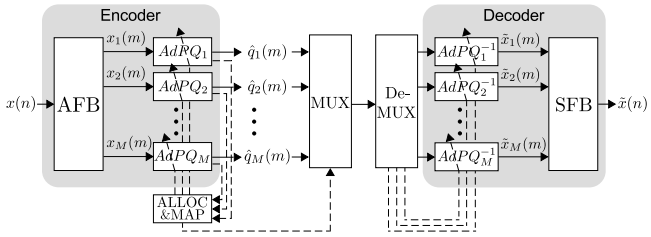

Fig. 1. Subband scheme for low delay audio coding extended by dynamic bit-allocation and mapping to preset.

The bit-allocation itself is done by means of a procedure explained in the following. Figure 2 shows a flow graph of the strategy. The inputs of the algorithm are the subband prediction error level estimates for each subband. In our modified version they are used for a dynamic initialization of the subband bit counters before entering the assignment loop. This loop contains three steps. First, a search for the band with the maximum level is performed. Second, the band i with the maximum level gets assigned an additional bit. In a third step the level value is reduced by a predefined value X_i. If there are bits left to assign, step one to three are repeated. Otherwise the assignment is terminated and in our modified version a mapping to the closest predefined preset is done.

In [7] the level estimate $s_i^2(n)$ used as input of the bit-allocation is the same that is used for the prediction error normalization and is calculated as variance estimate by

$$s_i^2(n) = (1-\alpha)\cdot s_i^2(n-1)+\alpha\cdot\tilde{e}_i^2(n) \text{ with } \alpha = 0.03125 \quad (1)$$

where $\tilde{e}_i^2(n)$ is the quantized prediction error of the i-th subband.

Using this level estimate in our coding scheme, we found that for a constant α finding the tradeoff between good estimation and fast adaption is basically impossible and a rather high α can lead to audible switching artifacts. Therefore we modified Formula 1 to contain a time variant α leading to

$$s_i(n) = (1-\alpha(n))\cdot s_i(n-1) + \alpha(n)\cdot v_i(n) \quad (2)$$

where

$$\alpha(n) = \begin{cases} \alpha_{AT} & \text{if } v_i(n) > s_i(n-1) \\ \alpha_{RT} & \text{else.} \end{cases} \quad (3)$$

and $\alpha_{AT} > \alpha_{RT}$ applies. In addition we use $v_i(n)$, which is the square root of the variance estimate used for prediction error normalization in our adaptive quantizer, instead of $\tilde{e}_i^2(n)$ for the subband level computation. Therefore, this level estimate can be interpreted as an adaptively smoothed version of the subband standard deviation estimates. This proved to give more consistent allocation results for a wide range of signal classes in our coding scheme.

As mentioned before, we also included a static and dynamic initialization of the bit-allocation. This means statically assigning a predefined number of bits to the lowpass and the first bandpass and allowing for a dynamic assignment of a predefined number of bits to the band with maximum level and the neighboring band before entering the allocation loop. The latter can be advantageous for signals that have rather high signal energy in the higher frequencies like percussive instruments or such with a lot of harmonics.

The level reduction applied to the buffer of band i which contains the maximum level was generalized and modified to allow for a nonlinear and iteration-dependent variation and can be expressed in dB as

$$X_i(\beta,\gamma,\delta,j) = 20\cdot\log_{10}\left(\frac{1}{\beta\cdot(2\gamma)^{\delta\cdot j}}\right) \quad (4)$$

with

$$\beta > 1,\ \gamma > 0.5,\ \delta \geq 0 \quad (5)$$

and a parameter j in the exponent that is increased with every bit assignment to the i-th band inside the allocation loop ($j = 0, 1, ...$). With this modification we intend to account for the fact that the contribution of every single bit to the subband SNRs does not necessarily match the 6 dB assumption used in [7], since it highly depends on the kind of quantization used and the shape of the probability density functions of the subband signals.

The parameters β, γ and δ are part of the global parameter optimization we present in Section 5. For making a computation of the exponentiation by a bitshift operation possible, γ and δ can be limited to be integers. Given $\delta = 0$ the level reduction X_i is the same in every iteration as proposed in [7].

While in [7] the allocation result is transmitted by three bits per subband, we modified the bit-allocation flow graph by adding a final post-processing step for limiting the signaling overhead. In this block, a mapping of the bit-allocation result to the closest one of a predefined and limited number of subband bit combinations is done. This procedure is based on our observation that for natural signals not all possible bit combinations appear and the frequency of occurrence

drops dramatically after the first ten to twenty most common ones. In addition we found that limiting the number of possible combinations has only a small or even a positive effect on the audio quality since it prevents unnecessary switches.

The mapping of a bit-allocation result vector $\mathbf{b}$ (ordered from the lowest band to the highest band) to the closest bit combination in the preset pool is done in two steps. First a comparison between the position of the biggest assigned value in the bit-allocation result and in all candidate presets is done. All presets that do not match in position are removed from the pool of candidate bit combinations. Then, for performing the final mapping, vector $\mathbf{b}$ is interpreted as a decimal number by calculating

$$D_{alloc.}(\mathbf{b}) = \sum_{i=1}^{M} \mathbf{b}(i) \cdot 10^{(M-i)} \tag{6}$$

where M is the number of subbands. This decimal representation is used in a search for the minimum absolute difference value between all preset combinations and the candidate allocation. The preset with the minimum difference is finally chosen and used for the next frame.

While this metric is straightforward it has proven to be sufficient for our task since it inherently incorporates the fact that lower bands are more important than higher ones. In addition it is also possible to efficiently implement the conversion to the decimal representation by approximating the multiplications with powers of ten by bitshift operations. As mentioned before with our mapping, the overhead for signaling of bit-allocations can be reduced significantly. With a limitation to 16 or 32 possible bit-allocations, which in our specific system usually is enough for the most common signal classes, the reduction therefore is in a range of up to 75%.

3. GLOBAL PARAMETER OPTIMIZATION

Given the fact that our bit-allocation scheme as well as the subband coding involve several partially interacting parameters that are not amenable to manual tuning, we were looking for a way of globally optimizing them with respect to psychoacoustic measures. Inspired by the PEAQ-based (Perceptual Evaluation of Audio Quality [10]) method presented in [11] we set up the framework shown in Figure 3. The core of this framework is an optimization algorithm that controls the parallel encoding and decoding of test set items and uses a PEAQ-based evaluation of the achieved audio quality for a given parameter vector $\mathbf{x}$. Like originally proposed in [11] the PEAQ results ODG_k for a given $\mathbf{x}$ are mapped to the cost function value by computing

$$C(\mathbf{x}) = \sum_{k=1}^{N} (\mathrm{ODG}_k(\mathbf{x}))^4 \tag{7}$$

where N is the number of test set items. This puts a stronger emphasis on signals with a worse audio quality without completely ignoring the results of better ones. We use a slightly modified version of the PEAQ c-code provided with [12] for calculation speedup. Since the global optimization has influence on the parameters of the bit-allocation, we also included a block which generates the pool of possible presets with a predefined size. It performs a histogram based pre-analysis of the bit-allocation distribution for a given parameter vector and returns the defined number of most common presets to the coding block.

In contrast to [11], the optimization itself is done by means of a genetic algorithm instead of a simulated annealing approach. This is because we found that for the genetic algorithm finding the tradeoff between guaranteed convergence and acceptable runtime without too much manual intervention is much easier. In addition it is well tested for directly solving mixed integer problems. For now we perform the optimization without doing a subsequent local search since we assume that the optimum found by global optimization is close enough to a local or the global minimum.

The parameters we optimize are:

- attack- and release-constants used in the adaptive quantizers of the subband processing
- parameters used for an amplitude scaling of the adaptive quantizer's static codebooks
- attack- and release-parameter (α_{AT}, α_{RT}) used in the bit allocation level estimation (Formula 2)
- number of bits used for initialization of the bit-allocation
- parameters β, γ and δ used for calculation of X_i in the assignment loop of the dynamic bit-allocation

and some other parameters of the error robust adaptive prediction and quantization that are omitted for brevity since they are not in the scope of this paper. Additional input parameters of the optimization are the desired bitrate, the number of presets for limiting the preset distribution and the frequency of bit-allocation calculation. The latter two determine the signaling overhead for the bit-allocation transmission.

The borders of the optimization are chosen to keep the codec in a "physically meaningful" operating point. The parameter values used for initialization of the optimization usually are the ones that resulted from a manual parameter adjustment during codec development.

4. EVALUATION AND RESULTS

For evaluation and demonstration of the influence of the dynamic bit-allocation and global optimization on the audio quality of our low delay audio coding scheme, we present PEAQ results for several test cases in the following.

The test items were taken from a database which is well known and often used for MPEG audio codec evaluation since it is a reasonable mixture of vocal and instrumental signals and contains some critical items for challenging codecs.

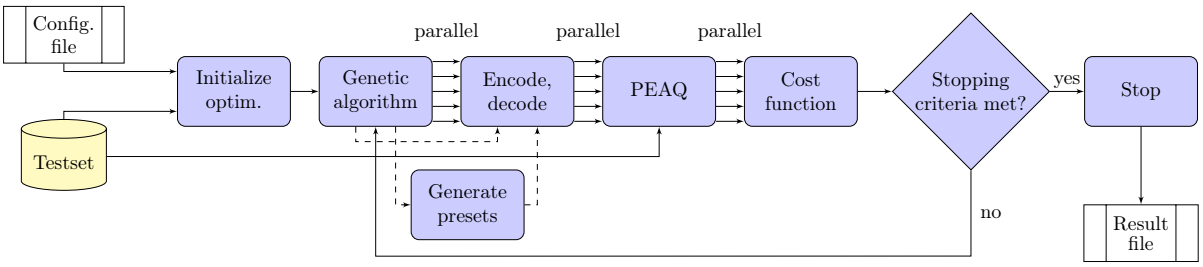

Fig. 3. Flow graph of the PEAQ-based global optimization using genetic algorithm.

Of course an optimization with this rather small database has a potential risk of overfitting. Since for now we are mainly interested in showing the general impact of the dynamic bit-allocation and parameter optimization on the audio quality of the codec, we skip a cross validation within this paper.

The evaluated test cases are:

- Case 1: no bit-allocation (fixed bit distrib. [6 5 3 3 3])
- Case 2: bit-allocation (manually adjusted parameters, without limited # of presets)
- Case 3: bit-allocation opt.+lim. (optimized parameters + limited # of presets)

The first case includes the constant bit allocation we used in [4] which, for ensuring a sufficient coding gain for wide-band signals, is a tradeoff between more bits for the lower and still enough bits within the higher bands. This results in a mean number of four bits per sample and therefore $176.4\,\mathrm{kbps}$ at $44.1\,\mathrm{kHz}$ sampling rate for the payload of the codec. For case two and three the data rate was set to correspond to the fixed allocation case while neglecting the small overhead for bit-allocation transmission.

Table 1 shows the PEAQ results for the different cases and test set items as well as the mean over all results for each case. Especially the ODG scores of the critical signals show the particular gain in audio quality that can be achieved by the use of the dynamic bit-allocation and the parameter optimization. On average the improvement from case 1 to case 3 is about 0.5 in the PEAQ score.

Of course a major part of the improvement in audio quality achieved by the dynamic bit-allocation is consequence to the fact that it enables for a temporary assignment of more that six bits in the lower bands. Nevertheless using a constant allocation with more than six bits in the lower bands and therefore less than three bits in the upper bands does not correspond to our goal of designing a wide-band audio codec.

	PEAQ ODG		
Item Name	Case 1	Case 2	Case 3
Vocal	-0.66	-0.69	-0.46
Male speech	-0.79	-0.74	-0.56
Female speech	-0.76	-0.63	-0.50
Trumpet	-1.16	-0.61	-0.43
Orchestra	-0.65	-0.71	-0.54
Big band	-0.42	-0.45	-0.40
Harpsichord	-0.46	-0.40	-0.32
Castanets	-2.08	-1.90	-0.93
Pitch pipe	-0.86	-0.47	-0.44
Bagpipe	-0.64	-0.51	-0.46
Glockenspiel	-3.25	-1.91	-0.82
Plucked strings	-0.49	-0.49	-0.51
Mean	-1.02	-0.79	-0.53

Table 1. PEAQ results for different test cases.

5. CONCLUSION

This paper presents a modification, extension and global optimization of a bit-allocation strategy for use in a low delay subband ADPCM-based audio coding scheme. The framework used for the assignment of bits to subband quantizers is based on a subband level estimation of the signals used for prediction error normalization.

It is modified to reduce switching artifacts by using a signal adaptive filter for subband level estimation. In addition it is extended to contain a mapping to predefined allocation presets which enables a reduction of signaling overhead.

Since our bit-allocation scheme as well as the subband coding involve several partially interacting parameters that are difficult to adjust manually, a framework for their global optimization is presented. It is based on a genetic algorithm using the PEAQ method for cost function calculation and is extended to include a unit for preset generation.

Experiments and their results show that in our coding scheme a significant improvement of audio quality can be achieved by the use of the modified bit-allocation. Compared to the reference method, the signaling overhead can be reduced by up to about 75%. Furthermore the global parameter optimization allows for an additional improvement of the mean PEAQ score of about 0.25 compared to manually adjusted parameters.

6. REFERENCES

[1] M. Schnell, M. Schmidt, M. Jander, T. Albert, R. Geiger, V. Ruoppila, P. Ekstrand, M. Lutzky, and B. Grill, "MPEG-4 Enhanced Low Delay AAC - a new standard for high quality communication," Oct. 2008, 125th AES Convention, San Francisco.

[2] U. Krämer, G. Schuller, S. Klier Wabnik, and J. Hirschfeld, "Ultra Low Delay audio coding with constant bit rate," Oct. 2004, 117th AES Convention, San Francisco.

[3] J.-M. Valin, T.B. Terriberry, C. Montgomery, and G. Maxwell, "A high-quality speech and audio codec with less than 10-ms delay," in *IEEE Transactions on Audio, Speech, and Language Processing*, 2010, number 1, pp. 58–67.

[4] S. Preihs and J. Ostermann, "Error Robust Low Delay Audio Coding based on Subband ADPCM," Oct. 2011, 131st AES Convention, New York.

[5] K. Nayebi, T. B. Barnwell, and M. J. T. Smith, "Low Delay FIR Filter Banks: Design and Evaluation," Jan. 1994, IEEE Transactions on Signal Processing, vol. 42, pp. 24–31.

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

[7] D. Martinez, F. Lopez, M. Rosa, and N. Ruiz, "A nearly transparent low delay audio coder," 1999, Recent Advances in Signal Processing and Communications (WSES Press).

[8] Allen Gersho and Robert M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, Norwell, MA, USA, 1991.

[9] Nuggehally S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*, Prentice Hall Professional Technical Reference, 1990.

[10] ITU Recommendation ITU-R BS.1287-1, *Method for objective measurements of perceived audio quality*, Nov. 2001.

[11] M. Holters and U. Zölzer, "Automatic parameter optimization for a perceptual audio codec," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2009*, Apr. 2009, pp. 13–16.

[12] P. Kabal, "An Examination and Interpretation of ITU-R BS.1387: Perceptual Evaluation of Audio Quality," Dec. 2003, Department of Electrical & Computer Engineering, McCill University, Montreal.