Embedded Audio Compression using Wavelets and improved Psychoacoustic Models

Markus Erne

Signal- and Information Processing Laboratory
Swiss Federal Institute of Technology ETHZ
erne@isi.ee.ethz.ch, http://www.isi.ee.ethz.ch

Abstract

In this paper, a new wavelet-based, scaleable approach to audio compression will be presented.
A signal-adaptive wavelet filterbank with an almost arbitrary time-frequency tiling allows to optimally adapt the filterbank in a rate-distortion-sense or depending on psychoacoustic criteria. No windowing or boundary wavelets are necessary in order to process the signal on a block-basis Based on the flexible time-frequency tiling of this new filterbank-approach, each subband can have an individual segmentation in time. The psychoacoustic model takes care of the control of the filterbank and performs an optimal decomposition of the audio signal. The scalability of the audio coder offers the ability to trade bitrate versus signal-quality and therefore graceful degradation can be achieved although the transmission bandwidth may be temporarily limited. An improved psychoacoustic model is integrated into the framework of the adaptive filterbank, the adaptive quantizer and the adaptive entropy-coding stage

1 Introduction

In the last few years, many high quality audio compression algorithms have been developed but most of them use a perceptual measure and operate at different but fixed bitrates. A variety of these algorithms are based on uniform polyphase filterbanks, modified discrete cosine transforms [1], using window switching or alternatively on lapped orthogonal transforms [2],[3]. Many proposals for wavelet-based audio coding schemes [4],[5] have been published recently. Uniform polyphase filterbanks can be implemented efficiently, but they do not well approximate the human auditory system and they do not offer large coding gains in a rate-distortion metric. Transform coders use block-based processing and show spectral distortion at the block-boundaries and pre-echo phenomena. The variety of existing musical instruments such as castanets, harpsichord or pitch-pipe exhibiting various coding requirements due to their completely different temporal and spectral fine-structure, suggests to use a filterbank with variable time-frequency resolution. Wavelet-filterbanks are known for a flexible time-frequency tiling but most wavelet-based audio coding algorithms are focussed to mimic the response of the human auditory system. It is the ultimate goal of this new audio coding algorithm to adapt the filterbank based on perceptual and rate-distortion criteria. Additionally, each subband can have a different segmentation in time, which allows to take care of the temporal fine-structure of the audio signal and to take advantage of temporal masking effects. A best basis search algorithm in a perceptual and rate-distortion sense for the wavelet-packet transform has been developed and implemented successfully.
Best-basis search algorithms in a rate distortion sense for wavelet-packet transforms have been published for a fixed time-segmentation [6] as "single-tree" algorithm as well as for variable time-segmentation over all subbands as "double-tree" algorithm [9]. We extended these techniques to a variable time-segmentation in every subband [7]. This framework allows to individually switch nodes of the wavelet-packet tree at completely different locations in time without affecting other nodes of the tree. The approach is well adapted to musical notation. In order to track each individual note, a flexible time-segmentation of every subband must be achieved and the position and the width of the subband in terms of pitch must be altered as well.

2 Adaptive Wavelet-filterbank

In order to optimize the coding gain based on perceptual and rate-distortion criteria, a flexible wavelet-filterbank has been developed. The filterbank offers a flexible time-frequency tiling and allows individual time-segmentation in every subband. In order to process finite length signals, a framework has been evaluated which enables the switching of the wavelet-packet bases in every node of the tree and additionally no blocking artifacts such as pre-echoes

or spectral distortion at the frame-boundaries can occur.

2.1 Boundary Conditions

The wavelet-packet transform can be written in matrix-form, using infinite matrices:

$$y_\infty = A_\infty x_\infty$$

In order to process finite length input vectors of length k, a submatrix A out of the infinite matrix A_∞ is selected in such a way that A has k columns and k-(N-2) rows for a given filterlength N. The decomposed signal y therefore can be written in matrix form: y=Ax. For the reconstruction, the synthesis matrix B is chosen to be a submatrix of A^T. B has k-(N-2) columns and k-2(N-2) rows. It can be shown [10] that k-2(N-2) samples out of the k samples can be reconstructed perfectly:

$$x_k = By_k$$

This framework allows to process overlapping blocks of input samples without using windowing or boundary wavelets. The algorithm has been extended to guarantee perfect reconstruction despite bases are switching from one node of the tree to another.

2.2 Switching Bases

By implementing a sophisticated memory management, a framework can be realized which allows to up- and down-switch the basis at every level of the tree. It is evident that for tree-levels near the root, the basis can be switched more frequently which matches to a requested high temporal resolution in the upper frequency bands whereas at the lowest tree-level with narrow subbands, a lower temporal switching-resolution due to fewer available samples can be tolerated.
All nodes of the wavelet-packet tree can be switched individually and the filterbank can fully adapt to the signal, depending on different criteria.

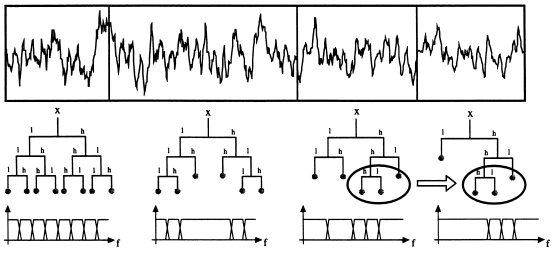

Figure 1: Individual switching of the wavelet bases, depending on the input signal

3 Best Basis Search

Having developed a framework for the individual switching of each node of the wavelet-packet tree, a measure on how to find the best basis for each signal interval has to be evaluated.

3.1 Rate-distortion measure

Best basis search algorithms have been published [6] and some of them make use of a least mean square error or a one-sided entropy metric. The momentary entropy in subband j at level i of the wavelet-packet tree is:

$$entropy_{ij}[k] = \frac{1}{N}\sum_{n=1}^{N} -\log_2(p_{ij}[k][quantized_{ij}[k-n+1]])$$

The reason we have chosen a common time measure for the up- and down-switching of every node now becomes obvious. In order to compare the entropy in every subband, we need to scale the entropy according to the number of samples in each subband. The scaled entropy in each subband is computed using a sliding window and a forgetting-factor for past samples before becoming part of a cost-function for each subband. The overall costs are compared for the parent node and both children nodes and depending on the result, the basis is switched up or down accordingly.
The same principle can be used if the scaled energy in every subband is used as a reference for switching the basis. Although the one-sided metrics such as entropy and energy do work well for fixed quantizers, they are not optimal in a rate-distortion sense. In [9], a method has been presented which jointly finds the optimal basis and the optimal quantization using the Langrangian cost function:

$$J(\lambda) = D + \lambda R$$

It can be shown that R-D optimality can be achieved when all leaves of the wavelet packet tree operate at a constant slope on their R-D curves. This approach will give best results in a rate-distortion sense, but it does not take any perceptual criteria into consideration.

3.2 Perceptual measure

For a perceptual measure, masking effects of the human auditory system are extremely important. In

frequency domain masking, a strong noise or a strong tone masker will mask the noise or a tone of the maskee [10].

All signals which are below the masking threshold will not be perceived by the human auditory system and therefore quantization noise in every subband can be as high as the masking threshold permits. In a subband coding system, every subband has an individual quantizer. It may be an advantage to have a subband decomposition equal to the critical bands of the human auditory system in order to profit of in-band masking. But again a flexible frequency tiling will enable to take care of inter-band masking (e.g. masking across critical bands).

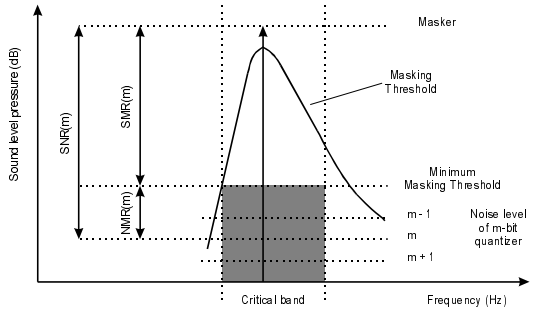

Figure 2: Frequency domain masking showing the masking threshold

Masking also occurs in the time domain. In the presence of abrupt signal transients, a listener will not perceive signals beneath the audibility threshold in the pre- and post-masking regions.

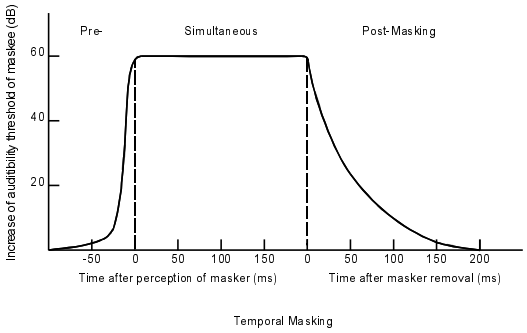

Figure 3: Temporal masking

Only a few available perceptual coders take advantage of temporal masking. A flexible, signal adaptive filterbank allows to analyze the temporal structure of the signal in every individual subband and to adapt the filterbank by taking profit of all masking effects in the frequency domain and in the time domain. With the current filterbank framework, pre-echoes can be avoided due to the individual segmentation in the time-domain for every subband. Similarly to the rate-distortion measure, we can define a perceptual measure for searching the best basis in a perceptual sense. A useful metric for estimating the achievable perceptual coding gain is based on perceptual entropy [11]. Perceptual entropy is therefore an estimate to the lower bound of transparent coding although it does not take into account rate-distortion criteria and temporal masking effects.

4 Weighted Cost Function

As it has been pointed out in the introduction, audio signals can have completely different temporal and spectral structure. Combining the rate-distortion measure and the perceptual measure in a weighted cost-function enables to cover applications such as lossless audio coding for archiving and audio-on-demand applications on the Internet with the very same coding scheme. Depending on the weight of the individual measures, the filterbank will adapt either in a rate-distortion sense or alternatively in a perceptual sense. Care has to be taken because these measures are not additive in terms of overall costs. The rate-distortion measure will operate in every subband but for the perceptual measure, a more global analysis in terms of frequency domain masking and temporal masking is used. An additional input to the cost-function is based on the complexity of the algorithm. As pointed out in [7], switching the basis will cause additional costs due to the redundant samples necessary for the reconstruction. If the complexity is to be kept as low as possible, switching the basis may be prohibited if the overall improvement in coding gain is rather small. Additionally, a "grid-function" for the switching can be set in order to avoid multiple up- and down-switching of the basis within a short segment of time.

5 Block diagram of the algorithm

The adaptive wavelet filterbank, the weighted cost function, including rate-distortion and perceptual measures have been presented in detail. The complete audio compression algorithm will include an adaptive quantizer and the entropy-coding stage. Both parts are under development because they mainly depend on the evaluations results and the performance of the adaptive filterbank and the cost-function metric. A control-loop enables to match the coding gain to the desired target bitrate and additionally a back-channel as proposed in the upcoming MPEG-4 standard can be used in order to guarantee graceful degradation in case of momentarily limited channel capacity.

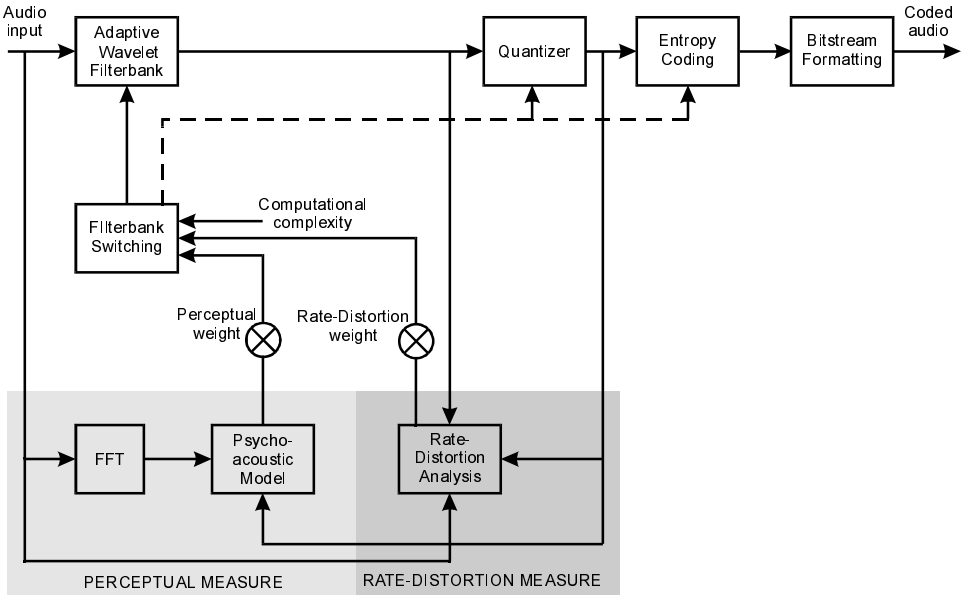

Figure 4: Overall block-diagram of the Wavelet-Audio Encoder

Additionally, advancement layers can be transmitted for scalable applications such as transmission over the Internet. Due to the flexible framework, this algorithm allows to scale quality down from lossles coding, using rate-distortion criteria only, down to to transmission over low-bandwidth channels, where the perceptual measure will have major impact.

6 Conclusions

A novel approach to a signal-adaptive filterbank for audio coding applications has been presented. In contrast to existing audio coding schemes, the algorithm allows individual time segmentation in every subband and every node of the wavelet-packet tree can be switched up and down in order to increase the coding gain. A weighted cost function allows to optimize the filterbank based on a perceptual or a rate-distortion measure. This system can perform lossless compression, near-lossless compression or perceptual compression of audio signals, depending on the weights which have been selected for the cost function. The cost-function additionally takes other parameters such as computational complexity and overall coding delay into consideration.

REFERENCES

[1] Brandenburg K., Stoll G., "The ISO/MPEG-Audio Codec: A Standard for Coding of High Quality Digital Audio", *AES Convention Preprint 3336,* March 1992

[2] Princen J., Johnston J.D., "Audio Coding with signal adaptive filterbanks", *Proceedings of ICASSP 95*, May 1995, pp. 3071-3073.

[3] Malvar H.S., "Signal Processing with Lapped Transforms", *Artech House,* Norwood, 1992.

[4] Sinha D., Tewfik A., "Low Bit Rate Transparent Audio Compression using Adapted Wavelets",. *IEEE Trans. on ASSP,* Vol. 41, No.12, December 1993, pp. 3463-3479.

[5] Kudumakis P, Sandler M., "On the Performance of Wavelets for Low Bit Rate Coding of Audio Signals",. *Proceedings of ICASSP 95*, May 1995, pp. 3087-3090.

[6] Wickerhauser M.V., "Adapted Wavelet Analysis from Theory to software", *IEEE Press,* 1994

[7] Faller C., Erne M., Moschytz G.S., "Wavelet Based Audio Compression", *Semesterarbeit an der ETH-Zürich,*, February 1998

[8] Coifman R., Wickerhauser M.V. "Entropy-Based Algorithms for Best Basis Selection" . *IEEE Trans. on Information Theory* Vol. 38, No. 2, March 1992 pp. 713-718

[9] Ramchandran K., Vetterli M., "Best Wavelet Packet Bases in a rate-Distortion Sense", *IEEE Trans. on Image Processing,* Vol. .2, No. 2, April 1993, pp. 160-175

[10] Zwicker E., Fastl H, ".Psychoacoustics Facts and Models" *Springer Verlag,* 1990

[11] Johnston J., "Estimation of Perceptual Entropy Using Noise Masking Criteria", *Proceedings of ICASSP 88*, May 1995, pp. 2524-2527