ADAPTIVE PRE- AND POST-FILTERING FOR A SUBBAND ADPCM-BASED LOW DELAY AUDIO CODEC

Stephan Preihs, Christoph Wacker and Jörn Ostermann

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

ABSTRACT

This paper addresses the combination of a low delay subband ADPCM-based audio codec with adaptive pre- and post-filtering for psychoacoustic noise shaping. We present how our basic scheme for error robust subband coding can be combined with two cascades of band shelving filters. The gain parameters of these filters are adapted by an algorithm that is based on power estimates which are obtained from the bandpass filtered subband signals of the codec.

Since the subband coding as well as the adaptive pre- and post-filtering include several partially interacting parameters that are hard to adjust manually, a framework for their psychoacoustic controlled global optimization is presented.

Experiments and their results show that the combination of our low delay subband ADPCM-based audio codec with adaptive pre- and post-filtering allows for a significant improvement of audio quality without any additional signaling overhead. Furthermore, the global optimization enables finding a psychoacoustical meaningful operating point.

Index Terms— Adaptive pre- and post-filtering, low delay audio coding, subband ADPCM, global parameter optimization

1. INTRODUCTION

When real time audio transmission systems have to deal with limited data rates, a demand for low bit rate low delay audio coding arises. If, in addition, a wireless audio transmission is supposed to take place in a live scenario with a high number of simultaneous channels, the constraints on delay and bandwidth become even more stringent. Furthermore, there are some applications that take place at the beginning of an "audio production chain" and therefore require a near transparent audio quality. This all together leads to a demand for low delay source coding algorithms with delays less than $1\,\mathrm{ms}$ and a very high audio quality.

Unfortunately the well known and established low delay codecs like AAC-ELD [1] and FhG ULD [2] or even the CELT (Opus) codec [3] are usually unsuitable for application in these scenarios since they either lead to a higher delay or do not provide an adequate audio quality. Nevertheless especially the FhG ULD [2] codec has proven that a combination of predictive coding with pre- and post-filtering can result in a sufficient audio quality.

Thus in [4] we presented a possible approach for audio coding with a very low delay of about $0.6\,\mathrm{ms}$ and near transparent audio quality. There, we combined a numerically optimized filter bank, designed with a variant of the framework published in [5], with an error robust version of the ADPCM presented in [6].

Figure 1 shows the basic structure of the codec (gray boxes). In the encoder an analysis filter bank (AFB) splits the input signal into $M = 5$ critically downsampled subband signals which are processed by an ADPCM-based time-domain coding. In the decoder the subband signals are decoded and a synthesis filter bank (SFB) 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.

Although 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 room for improvement. Therefore in [7] we extended our coding scheme by a signal adaptive dynamic bit-allocation and were able to show that with the modifications made a significant gain in audio quality can be achieved. As an alternative in this research study we raise the question if techniques for delay-free pre- and post-filtering can be adopted for use in our subband coding scheme for a psychoacoustic noise shaping that goes beyond the natural noise shaping of a subband codec.

2. ADAPTIVE PRE- AND POST-FILTERING

The main idea of pre- and post-filtering, which goes back to works like [8], is to post-filter the output of a lossy audio codec in such a way that the coding noise gets pushed towards or below the masking threshold. This means that the frequency response of the post-filter should nearly match this masking threshold and the pre-filter should have the exact inverse frequency response for not changing the frequency content of the codec's input signal.

While the pre- and post-filtering with exactly invertible minimum phase filters does not cause additional delay, the techniques presented in [8] make use of a block based psychoacoustic model that leads to an unavoidable delay of about 4-6 ms.

Therefore in [9] and following publications Holters et al. presented results on a delay-free audio coding scheme based on an ADPCM and show that adaptive pre- and post-filtering with a coarse estimation of the masking thresholds and a spectral shaping by dynamic compression in subbands is beneficial. Unfortunately their base ADPCM codec is not robust towards bit errors. Nevertheless, since the results published are quite promising, we modified their approach for use in our error robust subband coding scheme.

Figure 1 displays the combination of this pre- and post-filtering with our subband codec. Instead of using a block based psychoacoustic model, the necessary gains for the pre- and post-filter cascade of K bands are estimated directly from the reconstruction subband signals $\tilde{x}_j(m)$. This has the additional advantage that it does not require transmission of pre- and post-filter coefficients or, as for the dynamic bit-allocation approach, bit-allocation presets.

At this point we repeat the structure and formulation of the pre- and post-filter deployed by Holters et al. for being able to illustrate our modifications for use in the context of subband coding. For the

978-1-4799-7450-4/15/$31.00 ©2015 IEEE

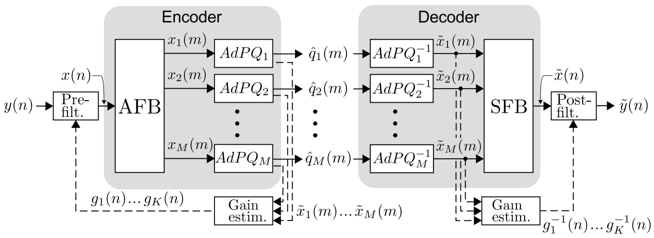

Figure 1: Flow graph of the subband codec structure (gray boxes) in combination with pre- and post-filtering and gain estimation.

derivation of the employed shelving filters from time continuous prototypes and further discussion about filter coefficient adaption the reader is referred to [9] and the references given there.

Figure 2 shows a flow graph of the encoder with adaptive pre-filtering. It can be divided into two main building blocks, the pre-filter cascade and the gain estimation, that will be described in the following. The post-filter in the decoder has the same structure but, for perfect inversion of the frequency response, the filters in the post-filter cascade are in the reverse order.

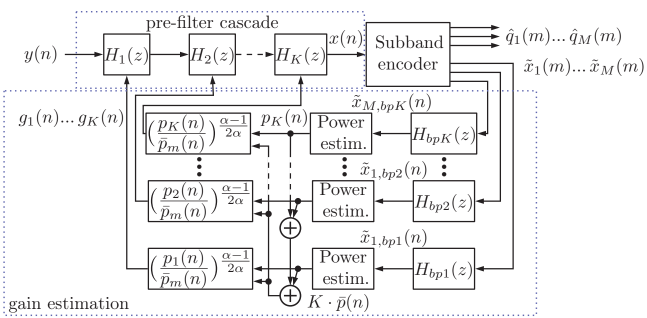

Figure 2: Flow graph of the subband encoder with gain estimation for cascaded adaptive pre- and post-filtering based on [9].

The first building block, which is the actual pre-filter, consists of a cascade of K shelving filters with the desired (and at first static) gain g_k in band k that result from first order "prototypes"

$$H_{k,prot.}(z) = \frac{C_k\sqrt{g_k} + 1 + (C_k\sqrt{g_k} - 1)\cdot z^{-1}}{\frac{C_k}{\sqrt{g_k}} + 1 + \left(\frac{C_k}{\sqrt{g_k}} - 1\right)\cdot z^{-1}}. \quad (1)$$

The coefficient C_k is introduced when the filters $H_{k,prot.}(z)$ are obtained from a time continuous prototype by bi-linear transformation for controlling the cutoff frequency and can be computed by

$$C_k = \tan\left(\frac{\omega_{w,k}}{2}\right) \quad (2)$$

where the width ω_w of band k is calculated from the desired upper and lower cutoff frequency by $\omega_{w,k} = \omega_{u,k} - \omega_{l,k}$ (ω denotes the normalized angular frequency within this paper).
The "prototypes" $H_{k,prot.}(z)$ are frequency shifted by replacing the unit delays with allpasses

$$A_k(z) = z^{-1}\,\frac{\cos(\omega_{c,k}) - z^{-1}}{1 - \cos(\omega_{c,k})z^{-1}} \quad (3)$$

where $\omega_{c,k}$ is the desired center frequency and can be computed by

$$\omega_{c,k} = 2\,\arctan\left(\sqrt{\tan\left(\frac{\omega_{l,k}}{2}\right)\cdot\tan\left(\frac{\omega_{u,k}}{2}\right)}\right). \quad (4)$$

This results in a set of second order shelving filters with the frequency responses

$$\left|H_k(e^{j\omega})\right| = \sqrt{\frac{(\cos(\omega_{c,k}) - \cos(\omega))^2 + (C_k\sin(\omega))^2 g_k}{(\cos(\omega_{c,k}) - \cos(\omega))^2 + \frac{(C_k\sin(\omega))^2}{g_k}}} \quad (5)$$

which can be implemented in a modified direct form filter for reducing the computational complexity resulting from the recalculation of the filter coefficient caused by the time variant gain $g_k(n)$.

The upper and lower cutoff frequencies $\omega_{u,k}$ and $\omega_{l,k}$ and therefore the center frequencies $\omega_{c,k}$ are chosen to nearly match the Bark-scale [10] with a lowest cutoff frequency of 0 Hz which results in $K = 25$. Figure 3 shows an example of the frequency responses $H_k(e^{j\Omega})$ (bottom) and the resulting pre-filter (top) when all g_k are set to $\sqrt{10}$ except for g_9 which is $0.1\cdot\sqrt{10}$. This leads to the notch at 1 kHz of the resulting pre-filter.

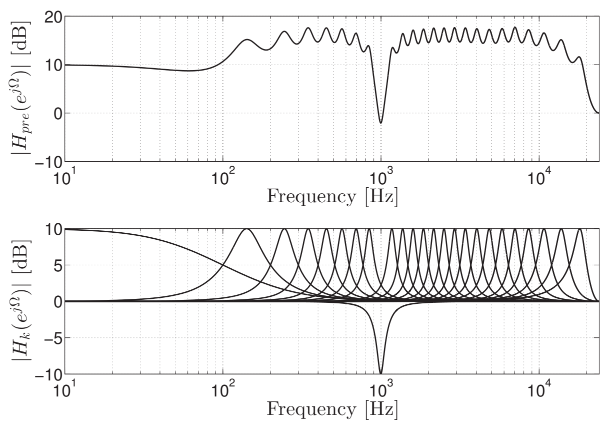

Figure 3: Frequency responses of second order shelving filters (bottom) and the resulting pre-filter (top) for all $g_k = \sqrt{10}$ except for g_9 which is $0.1\cdot\sqrt{10}$.

Due to the fact that, even after transformation, the shelving filters are only of second order, their frequency responses have a shape that looks more like a peak filter. This is why, for equal gain, the resulting pre-filter does not have the expected and desired flat frequency response. As already discussed by Holters et al. higher order filters would result in unwanted phase modulations of the shaped coding noise around the band edges and therefore the tradeoff between better band separation and resulting distortion lead to this low order. As we will show in Section 4, the non-ideal shape of the shelving filters does not prevent a sufficient noise shaping.

The second building block of the adaptive pre- and post-filtering is the gain estimation for dynamic compression which itself can be divided into the functional blocks bandpass filtering, power estimation and mapping to the final gain value. The power estimates $p_k(n)$ are calculated with the help of a first order recursive filter by

$$p_k(n) = (1-\lambda)p_k(n-1) + \lambda\left(\frac{\pi}{\omega_{w,k}}\tilde{x}^2_{j,bpk}(n) + p_{min}\right). \quad (6)$$

In our modification the input of this filter are the powers of the upsampled and bandpass filtered subband signals $\tilde{x}_{j,bpk}(n)$. They are normalized to the corresponding bandwidth. The bandpasses $H_{bpk}(z)$ used for filtering the upsampled subband reconstruction signals $\tilde{x}_j(n)$ result from a bi-linear transformation of a Butterworth prototype and are also frequency shifted by allpass transformation. The assignment of the subband reconstruction signals to the bandpasses is done such that the lower and upper cutoff frequency of bandpass $H_{bpk}(z)$ mostly lies within the frequency range of the analysis filter $H_{A_j}(z)$ of the analysis filter bank.

Equation 6 in addition contains a small parameters p_{min} that prevents the power estimate $p_k(n)$ from becoming zero. The sum of the band power estimates is further used for calculating the mean band power

$$\bar{p}(n) = \frac{1}{K}\sum_{k=1}^{K} p_k(n). \tag{7}$$

The final mapping to the signal adaptive gain values $g_k(n)$ for dynamic compression is done by normalization of $p_k(n)$ with $\bar{p}(n)$ and the use of an exponential curve

$$g_k(n) = \left(\frac{p_k(n)}{\bar{p}(n)}\right)^{\frac{\alpha-1}{2\alpha}} \tag{8}$$

with the factor $0 < \alpha \leq 1$ controlling the slope characteristics. Figure 4 shows examples of the exponential curve for different α.

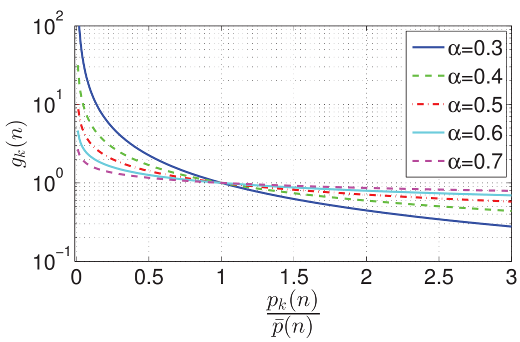

Figure 4: Examples of the exponential curve for mapping of normalized bandpass power estimates to gain $g_k(n)$ for different α.

The formulation of the shelving filters $H_k(z)$ allows for an exact inversion by just taking the inverse of the gain factor $g_k(n)$ for the adaption of the post-filter coefficients.

Discussion

As already discussed, the perfect inversion of the frequency response of the pre-filter is possible by applying the shelving filters of the post-filter in reverse order. Nevertheless the coefficient adaption can lead to nonlinear distortion if the smoothing parameter λ in the power estimation (Eq. 6) of the coefficient adaption is chosen too big. However, according to our experience, this distortion is rather uncritical for natural audio.

In contrast to Holters et al. we use a higher filter order for the bandpasses of the gain estimation than for the shelving filters of the pre- and post-filter since we experienced that this results in a higher consistency of the gain estimation. In addition we found that the, in our case unavoidable, delay between the input signal of the codec and the gain adaption of the pre-filter, which is caused by the delay of the analysis filter bank, is rather uncritical since the gain adaption is supposed to be slow anyway.

Of course the gain estimation could also be done on the basis of the codec's reconstruction signal $\tilde{x}(n)$ which is comparable to the approach presented by Holters et al.. Since this requires an extra synthesis filtering in the encoder and would lead to an additional lag in the gain adaption it is not preferable. As we will show in Section 4 the gain estimation from the subband signals does not have a negative impact on the pre- and post-filtering performance if compared to a gain estimation from a broadband signal.

In principle, with our robust subband coding scheme and the way of combining it with the adaptive pre- and post-filtering, the codec is robust towards bit errors occurring in the quantized and transmitted subband prediction errors $\hat{q}_j(m)$ since we can ensure the synchronization of the subband reconstruction signals $\tilde{x}_j(m)$ in encoder and decoder and there is no feedback loop in the gain calculation at the decoder side. Nevertheless evaluation of the bit error behavior goes beyond the scope of this paper.

3. GLOBAL PARAMETER OPTIMIZATION

Given the fact that the pre- and post-filtering 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 [11]) method presented in [12] we set up the framework shown in Figure 5. 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 [12] the PEAQ results ODG_i (Objective Difference Grade) for a given $\mathbf{x}$ are mapped to the cost function value $C(\mathbf{x})$ by computing

$$C(\mathbf{x}) = \sum_{i=1}^{N} (\mathrm{ODG}_i(\mathbf{x}))^4 \tag{9}$$

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 [13] for calculation speedup.

In contrast to [12], 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 and the shape properties of the adaptive quantizer's static codebooks,
- parameters α, λ and p_{min} of the gain estimation for the adaptive pre- and post-filtering

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. All the parameters remain fixed over time.

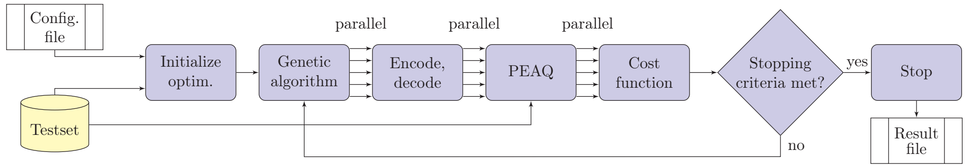

Figure 5: Flow graph of the PEAQ-based global optimization using genetic algorithm.

4. EVALUATION AND RESULTS

For evaluation and demonstration of the influence of the pre- and post-filtering 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 has often been 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. 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 pre- and post-filtering and parameter optimization on the audio quality of the codec, we skip a cross validation within this paper.
The test cases are (each case with one optimized parameter set):

- Case 1: No pre- and post-filtering,
- Case 2: Pre- and post-filtering (gain estimation from $x(n)$),
- Case 3: Pre- and post-filtering (gain estimation from $\tilde{x}_j(m)$),
- Case 4: Our dynamic bit-allocation approach from [7].

The first case includes the constant bit-allocation $(6, 5, 3, 3, 3)$ we used in [4] which, for ensuring a sufficient coding gain for wideband signals, is a tradeoff between more bits for the lower and still enough bits within the higher bands and therefore reflects our goal of designing a full band audio codec. This results in a mean number of four bits per sample and therefore e.g. $176.4\,\mathrm{kbps}$ at $44.1\,\mathrm{kHz}$ sampling rate for the payload of the codec.

Cases two and three are the codec using the same bit-allocation but in combination with the adaptive pre- and post-filtering for two different ways of gain estimation. The second case uses the pre-filtered signal $x(n)$ which is the input of the codec's analysis filter bank for gain estimation. Although not usable in the real coding system this serves as reference. This is because it is the closest we can get to the original gain estimation proposed by Holters et al. without additional synthesis filtering in the encoder and it is supposed to be the best we can get with this kind of gain estimation. The third case is our proposed gain estimation from the subband reconstruction signals $\tilde{x}_j(m)$. The fourth case are the results we presented for our globally optimized dynamic bit-allocation in [7].

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. Although for some items a small degradation in audio quality arises, the ODG scores of the critical signals Castanets, Glockenspiel and Trumpet show the particular gain in audio quality that can be achieved by the use of the pre- and post-filtering in combination with the parameter optimization. On average the improvement from case 1 to case 3 is about 0.15 in the PEAQ score. In addition there is hardly any difference between cases 2 and 3 which shows that our proposed gain estimation from the subband reconstruction signals is very similar to the one from a broad band signal. Optimizations with a different constant bit-allocation of $(6, 4, 4, 3, 3)$ showed even better results with an improvement of the mean PEAQ score from case one with 0.74 to case three with 0.43.

Of course the results presented in this section should be verified by a formal listening test like [14]. Unfortunately at the time of completion of this work this was not possible. Nevertheless the results were validated by informal listening experiments. One thing that was noticed during these tests is that, as expected, the performance of the noise masking resulting from the adaptive pre- and post-filtering approach clearly depends on the sound level used.

Table 1: PEAQ ODG results for different test cases and several test items as well as the mean over all values for each test case.

	PEAQ ODG			
Item Name	Case 1	Case 2	Case 3	Case 4
Vocal	-0.51	-0.52	-0.52	-0.46
Male speech	-0.56	-0.51	-0.56	-0.56
Female speech	-0.70	-0.58	-0.57	-0.50
Trumpet	**-0.99**	**-0.25**	**-0.37**	**-0.43**
Orchestra	-0.46	-0.26	-0.52	-0.54
Big band	-0.35	-0.27	-0.28	-0.40
Harpsichord	-0.31	-0.46	-0.44	-0.32
Castanets	**-1.32**	**-0.72**	**-0.76**	**-0.93**
Pitch Pipe	-0.48	-0.48	-0.49	-0.44
Bagpipe	-0.28	-0.23	-0.38	-0.46
Glockenspiel	**-1.71**	**-0.76**	**-0.74**	**-0.82**
Plucked Strings	-0.39	-0.60	-0.56	-0.51
Mean	-0.67	-0.47	-0.52	-0.53

5. CONCLUSION

In this paper we study the combination of an error robust low delay subband ADPCM-based audio codec with adaptive pre- and post-filtering for psychoacoustic noise shaping. We combine our basic scheme for error robust subband coding with two cascades of band shelving filters. Furthermore we show a way of adapting the gain parameters of these filters by the use of an algorithm that is based on power estimates which are obtained from bandpass filtered subband signals of the codec.

Since our error robust subband coding with its adaptive prediction and quantization as well as the adaptive pre- and post-filtering includes several partially interacting parameters that are hard to adjust manually, we present a framework for their psychoacoustic controlled global optimization.

Our experiments and their results show that the combination of a low delay subband ADPCM-based audio codec with adaptive pre- and post-filtering allows for a significant improvement of audio quality especially for critical signals without any additional signaling overhead. Furthermore, the global optimization enables finding a psychoacoustical meaningful operating point. This allows for a mean improvement of about 0.15 in the PEAQ ODG score.

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." 125th AES Convention, San Francisco, Oct. 2008.

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

[3] 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*, no. 1, pp. 58–67, 2010.

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

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

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

[7] S. Preihs and J. Ostermann, "Globally Optimized Dynamic Bit-Allocation Strategy for Subband ADPCM-Based Low Delay Audio Coding." 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Apr. 2015.

[8] G. Schuller, B. Yu, D. Huang, and B. Edler, "Perceptual audio coding using adaptive pre- and post-filters and lossless compression," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 6, pp. 379–390, Sept. 2002.

[9] M. Holters and U. Zölzer, "Delay-free lossy audio coding using shelving pre- and post-filters." IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Mar. 2008.

[10] E. Zwicker, "Subdivision of the audible frequency range into critical bands," *The Journal of the Acoustical Society of America, Volume 33, Number 2*, Feb. 1961.

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

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

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

[14] ITU-R Recommendation BS.1534-1, "Method for the subjective assessment of intermediate quality level of coding systems," Jan. 2003.