



Audio Engineering Society Convention Paper

Presented at the 116th Convention
2004 May 8–11 Berlin, Germany

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Low complexity parametric stereo coding

Erik Schuijers¹, Jeroen Breebaart², Heiko Purnhagen³, Jonas Engdegård³

¹*Philips Digital Systems Laboratories, Glaslaan 2, 5616 LW, Eindhoven, The Netherlands*

²*Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA, Eindhoven, The Netherlands*

³*Coding Technologies, Döbelnsgatan 64, 11352 Stockholm, Sweden*

Correspondence should be addressed to Erik Schuijers (erik.schuijers@philips.com)

ABSTRACT

Parametric stereo coding is a technique to efficiently code a stereo audio signal as a monaural signal plus a small amount of stereo parameters. The monaural signal can be encoded using any audio coder. The stereo parameters can be embedded in the ancillary part of the mono bit stream creating backwards mono compatibility. In the decoder, first the monaural signal is decoded after which the stereo signal is reconstructed from the stereo parameters. In this paper, a low complexity decoder solution is described based on complex-modulated filter banks. Combinations of the parametric stereo decoder with both a parametric coding scheme and with aacPlus will be elucidated.

1 INTRODUCTION

Parametric coding techniques gained major momentum in the field of audio coding. These techniques have been applied e.g. to create complete full bandwidth codecs [1], [2] as well as bandwidth extension algorithms like

Spectral Band Replication (SBR) [3]. Current developments in parametric audio coding focus on Parametric Stereo (PS) techniques [2], [4]. Using PS, a stereo signal is represented as a mono signal plus a small amount of parameters describing the stereo image. Recent re-

search on PS coding has led to a high quality stereo reconstruction at bit rates below 10 kbit/s for the stereo parameters [5]. However, as will be elucidated below, this system comes at high computational cost, especially in terms of memory usage. In this paper, a low complexity implementation is presented. Furthermore, combinations of this low complexity PS coding tool with a full bandwidth parametric audio coding scheme [2] and with aacPlus, the bandwidth-extended High Efficiency Advanced Audio Codec (HE-AAC) [6], are examined. The low complexity implementation of the PS system, as described in this paper, was proposed to MPEG-4 [7], [8] where the technical specification was finalized [9] and is awaiting formal approval.

The structure of this paper is as follows. Section 2 introduces the parametric stereo coding model. Section 3 presents a low complexity implementation of the parametric stereo decoding process. Combinations of this process with a parametric audio codec and with aacPlus are described in Sections 4 and 5, respectively. Finally, conclusions are drawn in Section 6.

2 PARAMETRIC STEREO CODING

Parametric Stereo coding aims at describing a stereo signal as a mono signal plus a set of parameters characterizing the stereo image. A block diagram of a PS encoder is shown in Figure 1. From the stereo input signal $(l[n], r[n])$, the time-variant stereo parameters are estimated on a non-uniform frequency grid, closely resembling the Equivalent Rectangular Bandwidth (ERB) grid [10]. These parameters describe the perceptually relevant spatial cues. Furthermore, a mono downmix $m[n]$ is generated. This mono downmix can then be encoded by any mono audio encoder. The stereo parameters are quantized and coded into the ancillary part of the mono bit stream yielding a backwards (mono) compatible system.

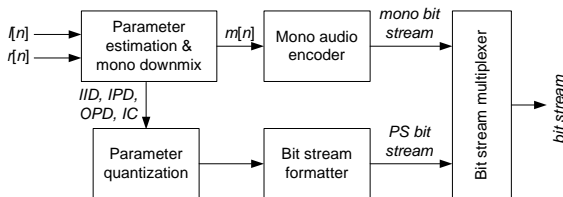


Figure 1: Generalized block diagram of PS encoder.

PS employs three types of parameters to describe the stereo image (see [5]):

1. Inter-channel Intensity Differences (IID); describing the intensity differences between the channels,
2. Inter-channel Phase Differences (IPD); describing the phase differences between the channels and
3. Inter-channel Coherence (IC); describing the coherence between the channels. The coherence is measured as the maximum of the cross-correlation as a function of time or phase.

In principle, these three parameters allow for a high quality reconstruction of the stereo image. However, the IPD parameters only specify the relative phase differences between the channels of the stereo input signal. They do not prescribe the distribution of these phase differences over the left and right channels. Hence, a fourth type of parameter is introduced, describing an overall phase offset or Overall Phase Difference (OPD).

In order to reconstruct the stereo image, in the PS decoder a number of operations are performed, consisting of scaling (IID), phase rotations (IPD/OPD) and decorrelation (IC). A block diagram of the PS decoder is shown in Figure 2.

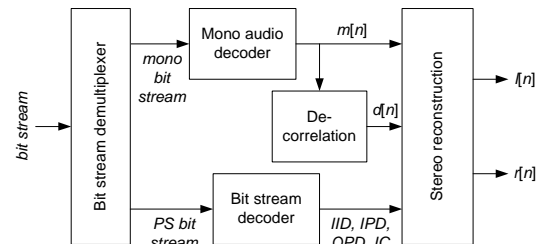


Figure 2: Generalized block diagram of PS decoder.

In the FFT-based PS decoder (see [2], [5]) first a decorrelated signal $d[n]$ is calculated by means of convolving the monaural signal $m[n]$ with a pre-defined sequence. In the stereo reconstruction process, consecutive windowed segments of both signals $m[n]$ and $d[n]$ are processed by a time-to-frequency (t/f) transform, performed by windowing followed by an FFT, resulting in the complex-valued frequency domain representations $M[k]$ and $D[k]$ respectively. The two frequency domain

representations of the left and right channels, $L[k]$ and $R[k]$ respectively, are obtained as linear combinations of the signals $M[k]$ and $D[k]$. The mixing parameters are time and frequency dependent; for each frequency component k the mixing process can be described by:

$$\begin{bmatrix} L[k] \\ R[k] \end{bmatrix} = \begin{bmatrix} h_{11}[k] & h_{12}[k] \\ h_{21}[k] & h_{22}[k] \end{bmatrix} \begin{bmatrix} M[k] \\ D[k] \end{bmatrix}, \quad (1)$$

where $h_{11}[k]$, $h_{12}[k]$, $h_{21}[k]$ and $h_{22}[k]$ are defined by the stereo parameters. The signals $L[k]$ and $R[k]$ are finally transformed back to the time domain by means of a frequency-to-time (f/t) transform. In the FFT-based PS decoder, the f/t transform consists of an inverse FFT followed by windowed overlap-add. For a more detailed overview of the actual parameter processing, we refer to [5].

3 LOW COMPLEXITY IMPLEMENTATION

For typical DSP-based applications like mobile devices, the computational complexity and memory usage of the decoder should be minimized in order to achieve e.g. maximum battery operation time. In the original FFT-based PS decoder [2], the complexity, both computationally as well as in terms of memory, is dominated by the time-to-frequency (t/f) and frequency-to-time (f/t) transforms that are applied [7]. This is primarily due to the length of the windows and FFT as they directly influence the length of input, output and intermediate storage buffers.

Recently, the SBR [3] tool for bandwidth extension of audio coding has been introduced. Similar to the PS coding paradigm, also SBR is a parametric audio coding enhancement tool that operates as post-processing in the decoder. Moreover, the structure of the SBR decoder and the PS decoder are quite similar. Both apply a t/f transform to obtain a frequency domain representation. In case of SBR, this is a band-limited representation, conveyed by an underlying audio coder like AAC, used to reconstruct a signal with full audio bandwidth. In case of PS, this is a mono representation used to reconstruct a stereo signal. Finally these representations are converted back to the time domain by means of an f/t transform.

The fact that both decoders are effectively post-processing algorithms means that a conventional decoder enhanced with such extensions is expanded with additional complexity. However, due to the fact that the

underlying conventional decoder operates at either a reduced sampling frequency in the case of SBR, or in mono in the case of PS, the additional complexity is at least partially compensated for.

The SBR algorithm makes use of complex-exponential modulated (Pseudo) Quadrature Mirror Filter (QMF) banks as t/f and f/t transforms, enabling flexible signal modification at high efficiency [11]. Therefore, they seem like a suitable alternative to the FFT employed in the decoder as presented in Section 2. Furthermore, the potentially very powerful combination of SBR with PS should not result in a decoder much exceeding the complexity of either SBR or PS. Hence, reuse of the t/f and f/t transforms of the SBR decoder is desirable.

3.1 Quadrature Mirror Filter Bank

In the analysis QMF bank, the complex-valued sub-band domain signals $s_k[n]$ are obtained as:

$$s_k[n] = \sum_{l=0}^{L-1} x[n-l]p[l]e^{j\frac{\pi}{K}(k+\frac{1}{2})(l+\phi)}, \quad (2)$$

where $x[n]$ represents the input signal, $p[n]$ represents the low-pass prototype filter impulse response of order $L-1$, ϕ represents a phase parameter, K represents the number of bands and k the sub-band index with $k=0, 1, \dots, K-1$. The magnitude responses of the first few lower frequency bands of the 64 bands analysis filter bank are illustrated in Figure 3.

The sub-band domain signals $s_k[n]$ are downsampled by a factor of K resulting in the downsampled complex sub-band domain signals $\zeta_k[n]$:

$$\zeta_k[n] = s_k[Kn]. \quad (3)$$

These downsampled sub-band domain signals can then be manipulated, e.g. by SBR or PS processing, resulting in the processed signals $\hat{\zeta}_k[n]$.

In the synthesis QMF bank, first the complex-valued sub-band domain signals $\hat{s}_k[n]$ are obtained by upsampling $\hat{\zeta}_k[n]$ with a factor of K :

$$\hat{s}_k[n] = \begin{cases} \hat{\zeta}_k[n/K] & \text{if } n = \dots, -2K, -K, 0, K, 2K, \dots \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

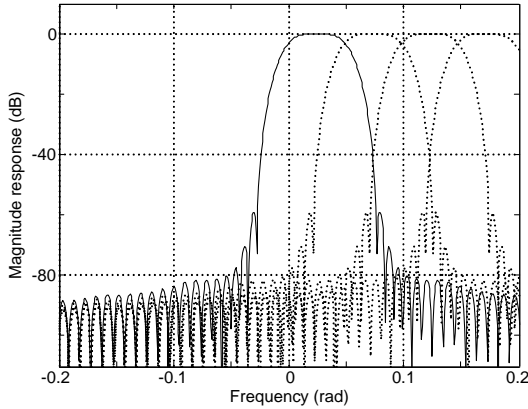


Figure 3: Magnitude responses of the 64 bands SBR complex-exponential modulated analysis filter bank for the first few lower frequency bands. The magnitude response for $k = 0$ is highlighted.

The reconstructed output signal $\hat{x}[n]$ is then obtained by:

$$\hat{x}[n] = 2\Re \left\{ \sum_{k=0}^{K-1} \sum_{l=0}^{L-1} \hat{s}_k[n-l] p[l] e^{-j\frac{\pi}{K}(k+\frac{1}{2})(l+\varphi)} \right\}, \quad (5)$$

with φ a phase parameter. Proper choice of constants and design of the prototype filter $p[n]$ results in a number of interesting properties:

1. Near-perfect reconstruction; if the downsampled sub-band domain signals $\zeta_k[n]$ are not modified, i.e., $\hat{\zeta}_k[n] = \zeta_k[n]$, the signal $\hat{x}[n]$ is a near-perfect reconstruction of $x[n]$.
2. Large stop-band attenuation; in order to provide a good discrimination between frequency bands, it is desired that adjacent frequency bands of the t/f and f/t transform only marginally influence each other.
3. Oversampled representation; typical transforms and filter banks employed in audio coding, like an MDCT, exhibit the so-called “critical sampling property,” obtained by means of aliasing cancellation. In order to prevent aliasing effects if the frequency domain representation of the signal is modified, the t/f transform must be oversampled.
4. Low memory requirements; the memory required for t/f and f/t transforms is mainly dependent on the

number of bands and the length of the prototype filter.

5. Computationally efficient; it is desired that the calculation of the t/f and f/t transform can be calculated by means of a computationally efficient algorithm, like e.g. the Discrete Fourier Transform can be calculated by means of an FFT.
6. Near analytical representation; except for the first and last sub-band, a near analytical representation is obtained. Hence, phase manipulations can be performed in a simple manner.

3.2 Hybrid filter bank for improved frequency resolution

For a typical sampling frequency of 44.1 kHz, the 64 bands analysis filter bank results in an effective bandwidth of approximately 344 Hz per band. However, there is considerable evidence that the spectral resolution of the (binaural) auditory system closely follows the ERB scale [12], [13]. This means that at low frequencies the binaural auditory system has a much finer resolution than the one given by the analysis filter bank as described above. In order to capture the perceptually relevant cues at a sufficient frequency resolution, the filter bank is extended. For the lower sub-bands, an additional sub-band filtering is carried out by means of oddly-modulated M^{th} band filter banks [14]. The analysis filtering for sub-band k is described by:

$$q_{k,m}[n] = \sum_{\lambda=0}^{\Lambda_k-1} \zeta_k[n-\lambda] g_k[\lambda] e^{j\frac{2\pi}{M_k}(m+\frac{1}{2})(\lambda-\frac{\Lambda_k-1}{2})}, \quad (6)$$

with Λ_k the the prototype filter length, $g_k[\lambda]$ the prototype filter, M_k the number of frequency bands, and $m = 0, 1, \dots, M_k - 1$ the frequency index of the resulting sub-sub-band signals $q_{k,m}[n]$.

Similarly to the signals $\zeta_k[n]$, the (sub-)sub-band signals $q_{k,m}[n]$ can be processed resulting in $\hat{q}_{k,m}[n]$. However, if we assume no processing, i.e., $\hat{q}_{k,m}[n] = q_{k,m}[n]$, and apply a synthesis operation described by:

$$\hat{s}_k[n] = \sum_{m=0}^{M_k-1} \hat{q}_{k,m}[n], \quad (7)$$

it can be shown that by using a prototype filter with the following constraints:

$$g_k[n] = \begin{cases} \frac{1}{M_k} & \text{if } n = \frac{\Lambda_k - 1}{2} \\ 0 & \text{if } n = \frac{\Lambda_k - 1}{2} + rM_k \end{cases} \quad (8)$$

where $r = \dots, -3, -2, -1, 1, 2, 3, \dots$, a perfectly reconstructing filter bank is formed by Equations 6 and 7.

For the sub-band signals that are not decomposed in separate sub-sub-bands, delay compensation is applied:

$$q_{k,0}[n] = s_k[n - \frac{\Lambda_k - 1}{2}], \quad (9)$$

3.3 Hybrid filter bank configurations

In the low complexity PS system two types of filter bank configurations have been defined. The first configuration is suited for 10 or 20 sets of stereo parameters. Since each parameter set represents a certain frequency range, this is also referred to as “10, 20 stereo band configuration.” Hybrid filtering is applied to the first 3 QMF bands with $M_0 = 8$, $M_1 = 4$, and $M_2 = 4$. For all $k = 3, \dots, 63$ delay compensation is applied according to Equation 9. In order to further reduce the complexity of this configuration, some of the filter bank outputs have been summed. For $k = 2, 3$ this leads to filters with a real-valued impulse response. As illustrated in Figure 4, this configuration results in a total of 71 (sub-)sub-bands.

The second filter bank configuration is suited for 34 sets of stereo parameters with $M_0 = 12$, $M_1 = 8$, $M_k = 4$ for $k = 2, 3, 4$. For all $k = 5, \dots, 63$ delay compensation is applied (see Equation 9). This results in a total of 91 (sub-)sub-bands.

To simplify time synchronization, the order $\Lambda_k - 1$ of the (sub-)sub-band filters have been chosen identical for all k , i.e., $\Lambda_k - 1 = 12$ for both filter bank configurations.

As an example the magnitude response of the 4 bands filter bank part of the 91 bands hybrid filter bank in sub-bands $k = 2, 3, 4$, is given in Figure 5. Obviously, due to the limited prototype length Λ_k , the stopband attenuation is only in the order of 20 dB.

3.4 De-correlation in the (sub-)sub-band domain

In the FFT-based PS decoder [2], [5] the de-correlated signal is obtained by means of convolution with a pre-defined de-correlation sequence in the time domain. In

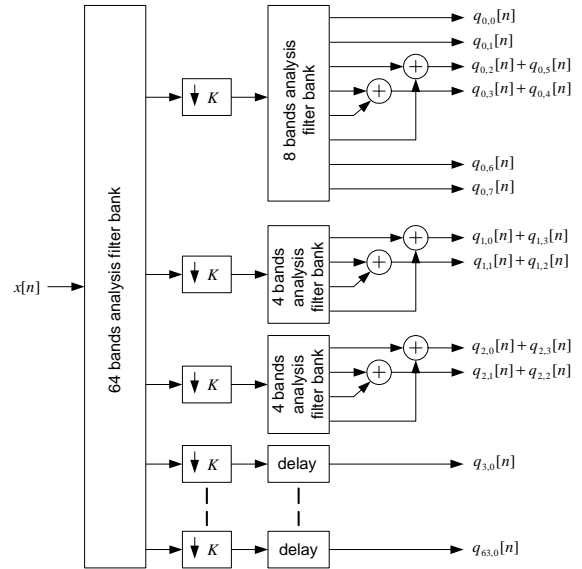


Figure 4: Block diagram of low complexity hybrid filter bank with $M_0 = 8$, $M_1 = 4$ and $M_2 = 4$, some of the filter bank outputs have been combined to reduce complexity.

order to obtain a perceptually satisfying result for the synthetic ambience, achieved with the de-correlated signal, a lengthy sequence needs to be employed. However, the QMF-based system allows for an alternative and more powerful approach for calculating the de-correlated signal. This method is based on applying reverberator-like circuits, or even simple delays, to the mono (sub-)sub-band domain signals. In this way, only one analysis t/f transform is required, thus reducing the overall complexity. For more details on this process, we refer to [15].

3.5 Stereo synthesis in (sub-)sub-band domain

Similar to the FFT-based decoder (see Eq. 1), the left and right (sub-)sub-band domain output signals, $L_n[k, m]$ and $R_n[k, m]$ respectively, are obtained by:

$$\begin{bmatrix} L_n[k, m] \\ R_n[k, m] \end{bmatrix} = \begin{bmatrix} h_{11,n}[k, m] & h_{12,n}[k, m] \\ h_{21,n}[k, m] & h_{22,n}[k, m] \end{bmatrix} \begin{bmatrix} M_n[k, m] \\ D_n[k, m] \end{bmatrix}, \quad (10)$$

where $n = 0, 1, \dots, N - 1$ with N the frame length. The matrices $h_{11,n}$, $h_{12,n}$, $h_{21,n}$, and $h_{22,n}$ are determined as

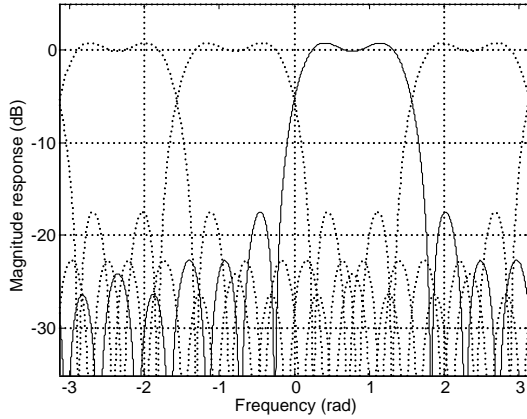


Figure 5: Magnitude response of the 4 bands filter bank, part of the 91 bands hybrid filter bank. The response for $m = 0$ has been highlighted.

follows. First, for each frame, the parameter positions n_i are extracted from the bit stream. For these parameter positions the vectors h_{11,n_i} , h_{12,n_i} , h_{21,n_i} and h_{22,n_i} are determined similar to the FFT-based decoder [5]. This is illustrated in Figure 6.

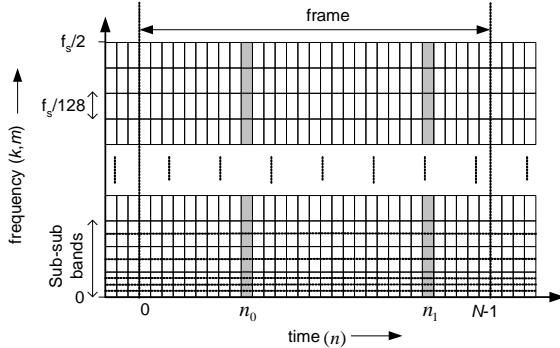


Figure 6: Time/frequency representation of (sub-)sub-band domain signals $q_{k,m}[n]$ for a frame ranging from $n = 0, 1, \dots, N-1$. Illustrated are two parameter positions n_0 and n_1 extracted from the bit stream.

For all $n \neq n_i$ the parameter manipulation matrices are calculated by means of linear interpolation:

$$h_n = (h_{n_i} - h_{n_{i-1}}) \frac{n - n_{i-1}}{n_i - n_{i-1}} + h_{n_{i-1}}, \quad (11)$$

for $n = n_{i-1} + 1, \dots, n_i - 1$ where the indices k and m have been discarded for clarity.

3.6 Overview of the resulting complete low complexity PS system

A block diagram of the resulting low complexity PS decoder is shown in Figure 7. First, the mono input signal $m[n]$ is transformed to the (sub-)sub-band domain signals $M_n[k, m]$ by means of the hybrid analysis filter banks as described above. These signals are used as input for the de-correlation process, resulting in the (sub-)sub-band domain signals $D_n[k, m]$. Both sets of (sub-)sub-band domain signals are used to reconstruct the stereo representation $L_n[k, m]$ and $R_n[k, m]$. Finally, these signals are transformed to the time domain by means of the hybrid synthesis filter banks resulting in the left and right output signals $l[n]$ and $r[n]$.

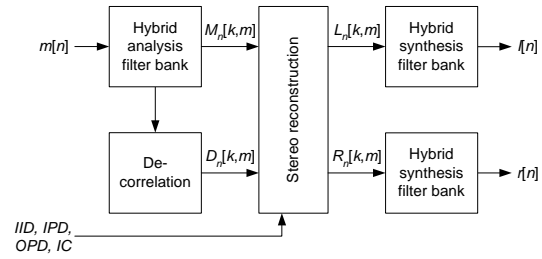


Figure 7: Block diagram of the low complexity PS decoder.

The resulting low complexity PS system allows for flexible configuration of the time and frequency resolution of the stereo parameters and supports different quantization accuracies. It is also possible to omit transmission of selected parameters completely. All this, in combination with time or frequency differential parameter coding and Huffman codebooks, makes it possible to operate this PS system over a large range of bit rates. For three typical configurations, Table 1 shows the average bit rates measured over a large set of stereo audio material. The corresponding quality levels of the stereo image range from medium to very high [5].

Table 2 shows complexity estimates for the FFT-based and the QMF-based PS decoder. It clearly shows that the complexity of the FFT-based system is dominated by the t/f and f/t transforms. The table also shows that a large complexity reduction has been obtained for the QMF-based decoder, especially in terms of RAM usage. The largest gain in complexity reduction is obtained by using alternative t/f and f/t transforms and the interpolation of the manipulation matrices.

Process	CPU [cycles/sample]			RAM [k words]		
	FFT	QMF		FFT	QMF	
		10, 20 band	34 band		10, 20 band	34 band
de-correlation	42	43	65	2.5	1.3	1.8
t/f and f/t transform	173	101	119	19.0	2.5	2.5
parameter processing	70	20	30	0.5	0.5	0.5
total	285	164	214	22.0	4.3	4.8
relative to FFT	100%	58%	75%	100%	20%	22%

Table 2: Estimated complexity of FFT-based and QMF-based PS decoder in both CPU cycles and RAM usage based on typical settings. For QMF-based decoder, both the 10, 20 and the 34 stereo band configurations are shown (taken from [7]).

	Configuration			Bit rate [kbit/s]
	# par.	IID quant.	IPD/OPD	
A	10	default	disabled	1.64
B	20	default	disabled	3.16
C	34	fine	enabled	9.02

Table 1: Measured average bit rates for the stereo parameters for three typical settings (44.1 kHz sampling rate, 23.2 ms time grid).

4 COMBINING PS WITH PARAMETRIC AUDIO CODING

Figure 8 gives a block diagram of the combination of PS with a parametric audio coding scheme, currently in the final standardization phase in MPEG-4 [9] as outlined in [2]. First, the bit stream is de-multiplexed and decoded into separate sets of parameters for transients, sinusoids, noise, and parametric stereo. The monaural PS input signal is generated by synthesis of the transient, sinusoid and noise parameters. Finally, the PS parameters are then used to reconstruct the stereo output signal.

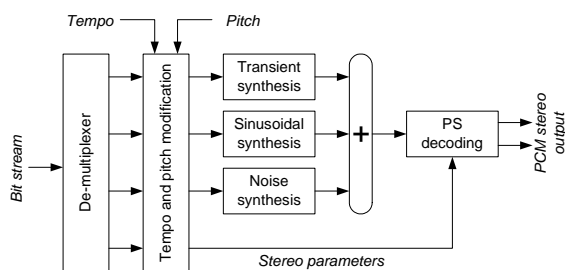


Figure 8: Block diagram of parametric audio decoder utilizing parametric stereo.

Figure 9 shows the Comparative Mean Opinion Score

(CMOS) [16] results of a comparative listening test of the parametric audio codec using a preliminary version of FFT-based PS system as described in [2] and the parametric audio codec using the QMF-based PS system as described in this paper. The mean gradings as well as the 95% confidence intervals are shown. The test was conducted employing nine subjects using headphones. The subjects were presented a reference signal, and two coded signals. They then had to grade codec A with respect to codec B with respect to the reference signal in scores ranging from +3 (Codec A much better than codec B) to -3 (Codec B much better than codec A). The scores show that at a lower computational complexity, the performance of the QMF-based PS system is better on average than the FFT-based system in a statistical sense. Furthermore, for none of the presented items, a statistically significant quality decrease is observed.

The low complexity PS system was proposed to MPEG-4 [7] and has successfully replaced the original FFT-based PS system [17] in the MPEG-4 high quality parametric audio coding scheme [9].

One particular advantage of parametric audio coding is the fact that typical post-processing algorithms like e.g. tempo and pitch modification can be performed with very little additional complexity in the parametric domain. As illustrated in Figure 8 prior to synthesis of the different parametric objects, the parameters are manipulated using a desired tempo and pitch modification.

In the original FFT-based parametric stereo decoder [2] both tempo and pitch modification of the stereo image can be performed fairly straightforward. Pitch modification can be applied by adjusting the frequency ranges to which each stereo parameter corresponds. Tempo modification of the PS object can be performed by adjusting the length of the PS analysis and synthesis windows.

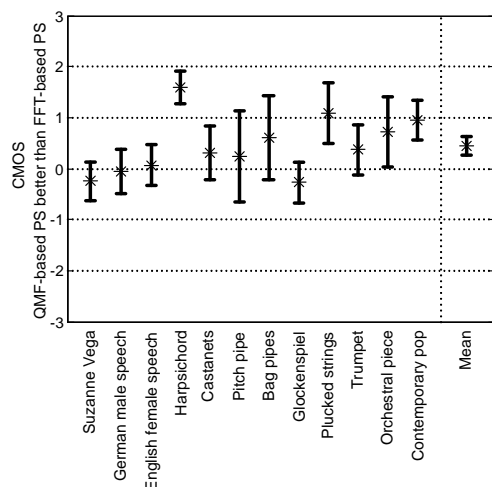


Figure 9: Listening test results comparing the parametric audio codec with the QMF-based parametric stereo coding scheme and the FFT-based parametric stereo coding scheme.

In the low complexity PS decoder pitch shifting can be applied in a similar way as in the FFT-based PS decoder. However, as a fixed time/frequency grid is employed, tempo modifications are not straightforward. Nevertheless, it is still possible to apply high quality tempo modification using the low complexity PS system by means of interpolation. This is illustrated in Figure 10. In case no time modification is applied, the parameter positions n_i are defined on the integer sub-band sample indices. For all $n \neq n_i$ interpolation is applied. This method can also be applied in case a tempo modification is applied. First the parameter positions n_i are scaled to αn_i . Then regular interpolation can be applied between these (not-necessarily integer) points, i.e., for $n = \lceil \alpha n_{i-1} \rceil, \dots, \lfloor \alpha n_i \rfloor$.

5 COMBINING PS WITH AACPLUS

The combination of MPEG-2/4 AAC with the SBR bandwidth extension tool is known as *aacPlus* and was standardized in MPEG-4 as the HE-AAC profile [18]. The basic principles of SBR have been elaborated on in several papers [3], [6], [11]. For the convenience of the reader a short review is given here.

The SBR principle stipulates that the missing high frequency region of a lowpass filtered signal can be recovered based on the existing lowpass signal and a small

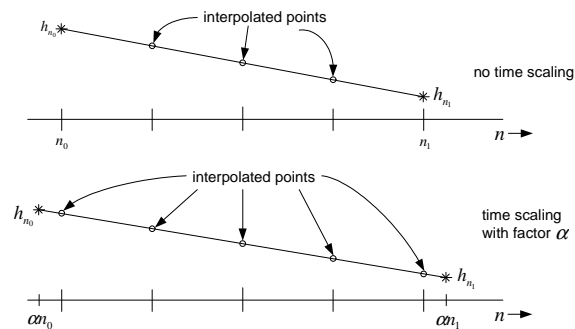


Figure 10: Time scaling of PS data using a fixed time/frequency grid. Top: interpolation of h_n in case of regular (non-time scaled) situation. Bottom: interpolation of h_n in case of time scaling by a factor of α .

amount of control data. The required control data is estimated in the encoder given the original wide-band signal. The *aacPlus* codec is a dual rate system, where the underlying AAC encoder/decoder is operated at half the sampling rate of the SBR encoder/decoder.

In the decoder, all SBR processing is done in the QMF domain. Hence, the output from the underlying AAC decoder is firstly analyzed with a 32 channel QMF filter bank. Secondly, the HF generator module recreates the highband by patching QMF sub-bands from the existing lowband to the highband. Furthermore inverse filtering is done on a per QMF sub-band basis, based on the control data obtained from the bit stream. The envelope adjuster modifies the spectral envelope of the regenerated highband, and adds additional components such as noise and sinusoids, all according to the control data in the bit stream. Since all operations are done in the QMF domain the final step of the decoder is a QMF synthesis to retain a time-domain signal. Given that the QMF analysis is done on 32 QMF sub-bands for 1024 time-domain samples, and the high frequency reconstruction results in 64 QMF sub-bands upon which the synthesis is done producing 2048 time domain samples, an up-sampling by a factor of two is obtained. Figure 11 a) shows the block diagram of an *aacPlus* decoder.

When the low complexity PS tool presented in this paper is combined with *aacPlus*, this results in a codec that achieves a significantly increased coding efficiency for stereo signals at very low bit rates when compared to *aacPlus* operating in normal stereo mode. Figure 11 b) shows a simplified block diagram of the resulting de-

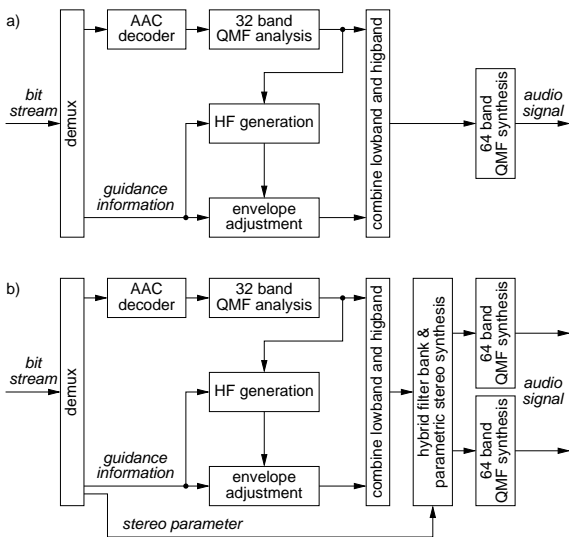


Figure 11: a) Block diagram of an *aacPlus* decoder. b) Block diagram of an enhanced *aacPlus* decoder utilizing parametric stereo.

coder, which is referred to as “enhanced *aacPlus*.” Since the SBR tool of *aacPlus* already operates in the QMF domain, the PS tool can be included in such a decoder in a computationally very efficient manner directly prior to the final QMF synthesis filter bank. Comparing Figures 11 a) and b), it is evident that only the parametric stereo decoding and synthesis, including its hybrid filter bank, have to be added to a mono *aacPlus* decoder, plus of course a second QMF synthesis filter bank. The computational complexity of such a decoder is approximately the same as that of a *aacPlus* decoder operating in normal stereo mode, where AAC decoding, QMF analysis filtering and SBR processing have to be carried out for both channels of a stereo signal. These complexity figures are based on the instrumentation of an optimized floating-point decoder implementation of the baseline version of the PS tool as defined in [9], which fully supports configurations like A and B in Table 1.

Figure 12 shows subjective results from a listening test comparing *aacPlus* using normal stereo coding at 24 and 32 kbit/s with enhanced *aacPlus* utilizing the parametric stereo tool at 24 kbit/s [8]. Two sites (indicated in black and gray) participated in this test, with 8 or 10 subjects per site, respectively. The 10 items from the MPEG-4 HE-AAC stereo verification test [19] were used as test material and playback was done using headphones. The

test employed MUSHRA [20] methodology and included a hidden reference and low-pass filtered anchors with 3.5 and 7 kHz bandwidth.

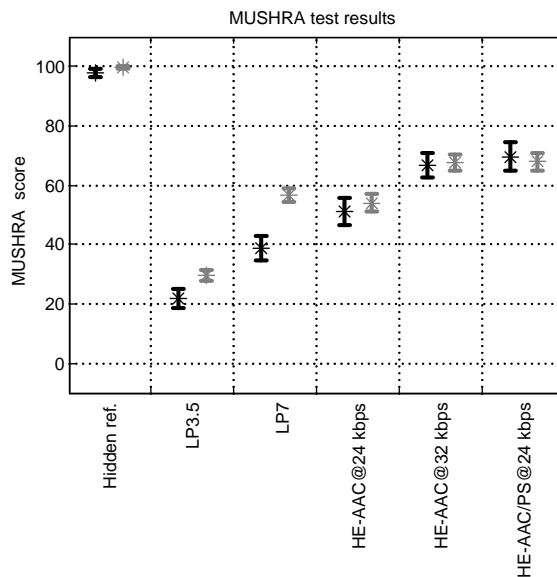


Figure 12: MUSHRA listening test results for two sites (black and gray) showing mean grading and 95% confidence interval (from [8]).

At both test sites, it was found that enhanced *aacPlus* with parametric stereo (HE-AAC/PS) at 24 kbit/s achieves an average subjective quality that is equal to *aacPlus* stereo (HE-AAC) at 32 kbit/s and that is significantly better than *aacPlus* stereo at 24 kbit/s. It is of interest to relate these results to the MPEG-4 verification test [19]. There, it was found that *aacPlus* stereo at 32 kbit/s achieved a subjective quality that was significantly better than AAC stereo at 48 kbit/s and was similar to or slightly worse than AAC stereo at 64 kbit/s. This shows that enhanced *aacPlus* achieves more than twice the coding efficiency of AAC for stereo signals. Further MUSHRA tests have shown that the enhanced *aacPlus* with parametric stereo achieves a significantly better subjective quality that normal *aacPlus* stereo also for 18 and 32 kbit/s.

The combination of the low complexity PS system with HE-AAC has also been adopted in MPEG-4 [8], [9].

6 CONCLUSIONS

A low complexity parametric stereo coding tool has been presented. It was shown that this parametric stereo coding tool significantly enhances the coding efficiency of existing audio coders. The presented tool is particularly interesting in combination with audio codecs using SBR bandwidth extension, since the resulting codec has approximately the same computational complexity as in a normal stereo configuration. The combination of AAC, SBR, and the parametric stereo tool presented here is referred to as enhanced aacPlus and enables coding of stereo signals at bit rates that are less than 50% of those required by AAC to achieve the same subjective quality.

7 REFERENCES

- [1] H. Purnhagen and N. Meine, "HILN – the MPEG-4 parametric audio coding tools," in *Proc. IEEE Int. Symposium on Circuits and Systems (ISCAS)*, Geneva, CH, May 2000, pp. III-201 – III-204.
- [2] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in *Proc. 114th AES Convention*, Amsterdam, The Netherlands, Mar. 2003, Preprint 5852.
- [3] M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral band replication, a novel approach in audio coding," in *Proc. 112th AES Convention*, Munich, Germany, May 2002, Preprint 5553.
- [4] C. Faller and F. Baumgarte, "Estimation of auditory spatial cues for binaural cue coding," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Orlando, Florida, USA, May 2002.
- [5] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "High-quality parametric spatial audio coding at low bitrates," in *Proc. 116th AES Convention*, Berlin, Germany, May 2004.
- [6] M. Wolters, K. Kjörling, D. Himm, and H. Purnhagen, "A closer look into MPEG-4 High Efficiency AAC," in *Proc. 115th AES Convention*, Los Angeles, USA, Oct. 2003, Preprint 5871.
- [7] W. Oomen, E. Schuijers, H. Purnhagen, and J. Engdegård, "MPEG4-Ext2: CE on low complexity parametric stereo," ISO/IEC JTC1/SC29/WG11 MPEG2003/M10366, Dec. 2003.
- [8] H. Purnhagen, J. Engdegård, W. Oomen, and E. Schuijers, "Combining low complexity parametric stereo with high efficiency AAC," ISO/IEC JTC1/SC29/WG11 MPEG2003/M10385, Dec. 2003.
- [9] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-3:2001/FDAM2 (parametric coding for high quality audio)," ISO/IEC JTC1/SC29/WG11 N6130, Dec. 2003.
- [10] B. Glasberg and B. Moore, "Derivation of auditory filter shapes from notched-noise data," in *Hearing Research*, 1990, vol. 47, pp. 103 – 138.
- [11] P. Ekstrand, "Bandwidth extension of audio signals by spectral band replication," in *Proc. 1st IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002)*, Leuven, Belgium, Nov. 2002, pp. 73 – 79.
- [12] J. Hall and M. Fernandes, "The role of monaural frequency selectivity in binaural analysis," in *J. Acoust. Soc. Amer.*, 1984, vol. 76, pp. 435 – 439.
- [13] A. Kohlrausch, "Auditory filter shape derived from binaural masking experiments," in *J. Acoust. Soc. Amer.*, 1988, vol. 84, pp. 573 – 583.
- [14] F. Mintzer, "On half-band, third-band and Nth band FIR filters and their design," in *IEEE Trans. Acoust., Speech, Signal Processing*, Oct. 1982, vol. ASSP-30, pp. 734 – 738.
- [15] J. Engdegård, H. Purnhagen, J. Rödén, and L. Liljeryd, "Synthetic ambience in parametric stereo coding," in *Proc. 116th AES Convention*, Berlin, Germany, May 2004.
- [16] ITU-T, "Methods for subjective determination of transmission quality," ITU-T Recommend. P.800, 1996.
- [17] ISO/IEC JTC1/SC29/WG11, "Text of ISO/IEC 14496-3:2001/FPDAM2 (parametric coding for high quality audio)," ISO/IEC JTC1/SC29/WG11 N5713, July 2003.
- [18] ISO/IEC, "Coding of audio-visual objects – Part 3: Audio, AMENDMENT 1: Bandwidth Extension," ISO/IEC Int. Std. 14496-3:2001/Amd.1:2003, 2003.

- [19] ISO/IEC JTC1/SC29/WG11, “Report on the verification tests of MPEG-4 High Efficiency AAC,” ISO/IEC JTC1/SC29/WG11 N6009, Oct. 2003.
- [20] ITU-R, “Method for the subjective assessment of intermediate quality level of coding systems (MUSHRA),” ITU-R Recommend. BS.1534, 2001.