LOW COMPLEXITY PARAMETRIC STEREO CODING IN MPEG-4

Heiko Purnhagen

Coding Technologies
Döbelnsgatan 64, 11352 Stockholm, Sweden
hp@codingtechnologies.com

ABSTRACT

Parametric stereo coding in combination with a State-of-the-Art coder for the underlying monaural audio signal results in the most efcient coding scheme for stereo signals at very low bit rates available today. This paper reviews those aspects of the parametric stereo paradigm that are important for audio coding applications. A complete parametric stereo coding system is presented, which was recently standardized in MPEG-4 Audio. Using complex modulated lter banks, it allows implementation with low computational complexity. The system is backward compatible and enables high quality stereo coding at total bit rate of 24 kbit/s when used in combination with aacPlus.

1. INTRODUCTION

The performance of low bit rate audio coding systems can be signicantly improved for stereo signals when a parametric stereo (PS) coding tool is employed. In such a system, a mono signal is conveyed using a State-of-the-Art audio coder and stereo parameters are estimated in the encoder and added as side information to the bit stream. In the decoder, the stereo signal is reconstructed from the decoded mono signal with help of the stereo parameters.

Techniques for joint stereo coding [1, Section 11.2.8], like intensity stereo (IS) and mid/side (M/S) coding, have long been used in audio coding systems. Intensity stereo coding can be seen as a simple form of parametric stereo, where the lateral localization of a mono signal is controlled by pan parameters. Because IS reuses the time-to-frequency mapping of the underlying audio coder to achieve frequency selective processing, it is prone to aliasing artifacts and typically only used for higher frequency bands. Furthermore, IS is unable to reconstruct the stereophonic ambience that might have been present in the original signal.

These and other aspects have been addressed recently by research on more general parametric stereo systems [2], [3], [4]. The aliasing problem can be overcome by using an oversampled time-to-frequency mapping like a complex-valued transform or lter bank. Stereo ambience can be reconstructed using appropriate techniques to decorrelate the two channels of a stereo signal in the decoder. Further research studied reconstruction of the time or phase differences between the stereo channels.

This paper reviews the progress in parametric stereo coding, considering the PS tool recently standardized in Extension 2 of MPEG-4 Audio [5] as an example. This PS tool combines low computational complexity with advanced means for recreating stereo ambience in the decoder. To assess the performance of a complete parametric stereo coding system, the combination of the PS tool with aacPlus will be studied here. The aacPlus coder, which is used to convey the full bandwidth mono signal is this system, is

the combination of Spectral Band Replication (SBR) [6] and Advanced Audio Coding (AAC) [7] and was standardized as High-Ef cienc y AAC (HE-AAC) in Extension 1 of MPEG-4 Audio [8].

This paper is organized as follows. Section 2 reviews the parametric stereo coding paradigm and presents the fundamental concepts of stereo parameter estimation in the encoder and stereo reconstruction in the decoder. Section 3 studies a low complexity implementation of a parametric stereo system. Optimized techniques to generate synthetic ambience are discussed in Section 4. Section 5 presents the combination of aacPlus and PS in MPEG-4 and reports the results of listening tests. Finally, conclusions are drawn in Section 6.

2. THE PARAMETRIC STEREO CODING PARADIGM

Parametric stereo coding is a technique to ef ciently code a stereo audio signal as a monaural signal plus a small amount side information for stereo parameters. The monaural signal can be encoded using any audio coder. The stereo parameters can be embedded in the auxiliary part of the mono bit stream, thus achieving full forward and backward compatibility. In the decoder, rst the monaural signal is decoded after which the stereo signal is reconstructed with help of the stereo parameters. Figures 1 and 2 show the generalized block diagram of an encoder and decoder, respectively, that employ parametric stereo coding.
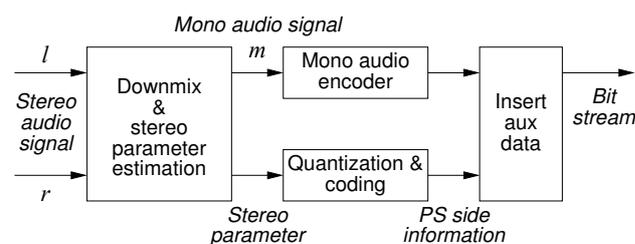

Figure 1: *Generalized block diagram of PS encoder.*

Three types of parameters can be employed in a parametric stereo system to describe the stereo image [9], [10], [11].

- Inter-channel Intensity Difference (IID), describing the intensity difference between the channels.

- Inter-channel Cross-Correlation (ICC), describing the cross-correlation or coherence between the channels. The coherence is measured as the maximum of the cross-correlation as a function of time or phase.

- Inter-channel Phase Difference (IPD), describing the phase difference between the channels. This can be augmented by
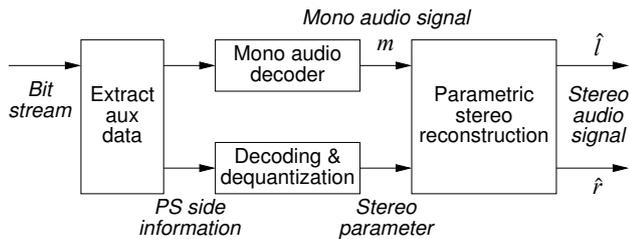
Figure 2: *Generalized block diagram of PS decoder.*

an additional Overall Phase Difference (OPD) parameter, describing how the phase difference is distributed between the channels. The Inter-channel Time Difference (ITD) can be considered as an alternative to IPD.

These stereo parameters vary over time and frequency. Psychoacoustics indicate that a Bark or ERB like frequency scale is appropriate for stereo parameters. Hence, in the PS systems discussed here, the audio bandwidth of 20 kHz is non-uniformly divided into 10, 20, or 34 stereo bands according to such a perceptual frequency scale. The temporal resolution of the stereo parameters is in the order of 10 to 50 ms.

To enable frequency-selective stereo analysis in the decoder and stereo reconstruction in the decoder, respectively, an appropriate time-to-frequency mapping is needed. To avoid artifacts caused by signal modification in the decoder, the frequency domain representation should be oversampled, and a complex-valued representation simplies time- or phase-related analysis and modification. Both lter bank and transform based approaches have been used successfully.

The remainder of this paper focuses on a system using only IID and ICC parameters. It is the baseline subset of the generalized parametric stereo tool (including IPD/OPD parameters) developed by Philips and Coding Technologies that was recently standardized in Extension 2 of MPEG-4 Audio [12], [13], [5].

2.1. Encoder

In order to enable time and frequency selective analysis and processing, the channels $l[n], r[n]$ of the stereo input signal are transformed into an oversampled complex frequency domain representation $L[k,i], R[k,i]$. Estimation of stereo parameters is carried out individually for each tile in the time-frequency plane. Such a tile is dened as the intersection of a short signal segment (frame) with $i \in I$ and a frequency region (stereo band) b with $k \in K_b$. For each tile, the IID and ICC parameters are estimated as follows.

$$\text{IID}[b] = 10\log_{10} \frac{\sum_{k \in K_b} \sum_{i \in I} L^*[k,i]L[k,i]}{\sum_{k \in K_b} \sum_{i \in I} R^*[k,i]R[k,i]} \tag{1}$$

$$\text{ICC}[b] = \frac{\text{Re}\left(\sum_{k \in K_b} \sum_{i \in I} L^*[k,i]R[k,i]\right)}{\sqrt{\left(\sum_{k \in K_b} \sum_{i \in I} L^*[k,i]L[k,i]\right)\left(\sum_{k \in K_b} \sum_{i \in I} R^*[k,i]R[k,i]\right)}} \tag{2}$$

A suitable mono downmix is obtained as a linear combination of both input signals.

$$M[k,i] = g_l L[k,i] + g_r R[k,i] \tag{3}$$

where g_l and g_r are appropriate weights. With $g_l = g_r = 1/2$ a normal mono downmix is obtained.

2.2. Decoder

Based on the decoded mono signal m, both channels $\hat{l}, \hat{r}$ of the stereo signal are reconstructed with help of the stereo parameters *IID* and *ICC* in the decoder. Also for this stereo synthesis process, an oversampled complex frequency domain representation $M[k,i], \hat{L}[k,i], \hat{R}[k,i]$ is employed and the same tiling of the time-frequency plane as in the encoder is used.

In order to be able to reconstruct stereo ambience as characterized by the ICC parameter, a decorrelated version D of the decoded mono signal M with $\text{E}\{M^*D\} \approx 0$ is generated by means of an appropriate all-pass lter (see Section 4). The decorrelated signal has (approximately) the same spectral and temporal energy distribution as the mono signal, i.e., $\text{E}\{D^*D\} = \text{E}\{M^*M\}$.

Based on $M[k,i]$ and $D[k,i]$, the stereo signal $\hat{L}[k,i], \hat{R}[k,i]$ can now be reconstructed as

$$\begin{bmatrix} \hat{L}[k,i] \\ \hat{R}[k,i] \end{bmatrix} = \mathbf{H}[k,i] \begin{bmatrix} M[k,i] \\ D[k,i] \end{bmatrix} \tag{4}$$

using the up-mix matrix $\mathbf{H}$.

$$\mathbf{H} = \begin{bmatrix} c_l \cos(\beta + \alpha) & c_l \sin(\beta + \alpha) \\ c_r \cos(\beta - \alpha) & c_r \sin(\beta - \alpha) \end{bmatrix} \tag{5}$$

with $c = 10^{\text{IID}/20}$, $c_l = \sqrt{2}c/\sqrt{1+c^2}$, $c_r = \sqrt{2}/\sqrt{1+c^2}$, and $\alpha = \arccos(\text{ICC})/2$ reconstructs $\hat{L}, \hat{R}$ such that they fulll

$$\text{E}\{\hat{L}^*\hat{L}\} + \text{E}\{\hat{R}^*\hat{R}\} = 2\text{E}\{M^*M\} \tag{6}$$

and the spatial characteristics as described by the stereo parameters IID and ICC estimated in the encoder. To nd an appropriate value for the remaining free parameter β, the amount of the (undecorrelated) mono signal M in $\hat{L} + \hat{R}$ is maximized, which gives

$$\beta = \arctan\left(\tan(\alpha)\frac{c_r - c_l}{c_r + c_l}\right). \tag{7}$$

This process can be considered as a rotation in the M, D-plane and is shown in Figure 3. Equation 7 can be simplied using the approximation

$$\beta = \alpha \frac{c_r - c_l}{\sqrt{2}}. \tag{8}$$

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 an earlier FFT-based version [3] of the PS decoder considered here, 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 [12].
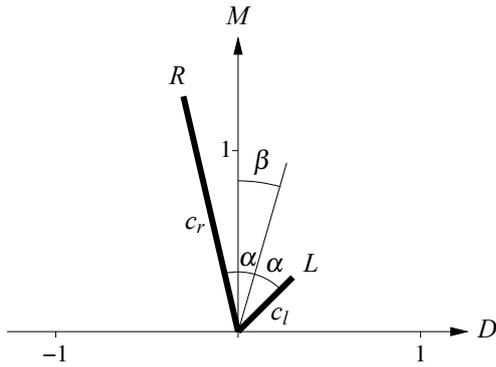
Figure 3: *Rotation for IID = -10 dB and ICC = 0.5 showing how L and R are composed from M and D by means of* **H***.*

Recently, the SBR tool [6] for bandwidth extension of audio coding has been introduced. Similar to the PS tool, also SBR is a parametric audio coding enhancement tool that operates as post-processing in the decoder. Moreover, the structure of the SBR tool and the PS tool in the decoder are similar. Both first apply a t/f transform to obtain a frequency domain representation, then carry out processing in this domain, and finally apply an f/t transform to convert the processed signal back to the time domain.

The fact that both tools are post-processing algorithms means their complexity adds to that of the underlying audio decoder. However, due to the fact that the underlying decoder operates at either a reduced sampling rate in the case of SBR, or in mono in the case of PS, it is less complex that in conventional full-bandwidth or stereo operation. This effect compensates for most of the complexity added by the SBR or PS tools.

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 with high computational efficiency [14]. Therefore, it is a suitable alternative to an FFT-based implementation of a PS decoder.

3.1. Quadrature Mirror Filter Bank

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

$$s_k[n] = \sum_{l=0}^{L-1} x[n-l]p[l]e^{j\frac{\pi}{K}\left(k+\frac{1}{2}\right)(l+\phi)} \qquad (9)$$

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 $k = 0, 1, ..., K-1$. The sub-band domain signals $s_k[n]$ are downsampled by a factor of K resulting in the downsampled complex sub-band domain signals $\varsigma_k[n] = s_k[Kn]$.

These downsampled sub-band domain signals can then be manipulated, e.g. by SBR or PS processing, resulting in the processed signals $\hat{\varsigma}_k[n]$. In the synthesis QMF bank, first the complex-valued sub-band domain signals $\hat{s}_k[n]$ are obtained by upsampling $\hat{\varsigma}_k[n]$ with a factor of K. Then, the reconstructed output signal $\hat{x}[n]$ is then obtained as

$$\hat{x}[n] = 2\,\mathrm{Re}\left\{ \sum_{k=0}^{K-1}\sum_{l=0}^{L-1} \hat{s}_k[n-l]p[l]e^{-j\frac{\pi}{K}\left(k+\frac{1}{2}\right)(l+\varphi)} \right\} \qquad (10)$$

with φ a phase parameter. Proper choice of constants and design of the prototype filter $p[n]$ results in a system with near-perfect reconstruction and large stop-band attenuation. The complex-valued sub-band signals are oversampled by factor 2, enabling aliasing-free signal manipulation, and are approximately analytic signals (except for the first and last sub-band).

3.2. Hybrid filter bank for improved frequency resolution

For a typical sampling rate of 48 kHz, the 64 band QMF bank results in an effective bandwidth of 375 Hz. However, a PS system with, e.g., 20 stereo bands following a perceptual frequency scale, asks for a finer frequency resolution than 375 Hz at low frequencies.

In order to capture the perceptually relevant cues at a sufficient frequency resolution, the QMF bank is extended. For the lower sub-bands, an additional sub-band filtering is carried out by means of oddly-modulated M band filter banks. The analysis filtering for sub-band k is described by

$$q_{k,m}[n] = \sum_{\lambda=0}^{\Lambda_k-1} \varsigma_k[n-\lambda]g_k[\lambda]e^{j\frac{2\pi}{M_k}\left(m+\frac{1}{2}\right)\left(\lambda - \frac{\Lambda_k-1}{2}\right)} \qquad (11)$$

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

Similarly to the signals $\varsigma_k[n]$, the sub-sub-band signals $q_{k,m}[n]$ can be processed resulting in $\hat{q}_{k,m}[n]$. Using the simple synthesis operation

$$\hat{s}_k[n] = \sum_{m=0}^{M_k-1} \hat{q}_{k,m}[n] \qquad (12)$$

this additional sub-band filtering achieves perfect reconstruction [10]. Also the complex-valued sub-sub-band signals $q_{k,m}[n]$ are approximately analytic signals.

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

$$q_{k,0}[n] = s_k[n - \frac{\Lambda_k - 1}{2}]. \qquad (13)$$

This combination of the QMF bank with additional sub-band filtering results in a hybrid filter bank. The PS system considered here typically uses 20 stereo bands and applies additional sub-band filtering to the first 3 QMF bands with $M_0 = 8$, $M_1 = 4$, and $M_2 = 4$. For all $k = 3, \ldots, 63$ delay compensation is applied according to Equation 13. 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. An alternative hybrid filter bank configuration, suited for 34 stereo bands, employs $M_0 = 12$, $M_1 = 8$, and $M_k = 4$ for $k = 2, 3, 4$ [10].

3.3. Stereo synthesis in (sub-)sub-band domain

In the decoder, first a (sub-)sub-band representation of the decoded mono signal *m* is obtained by means of the hybrid analysis filter bank presented above. The decorrelated signal *d* generated directly in the (sub-)sub-band domain as described in Section 4. Then, the (sub-)sub-band representation of both channels of the stereo signal
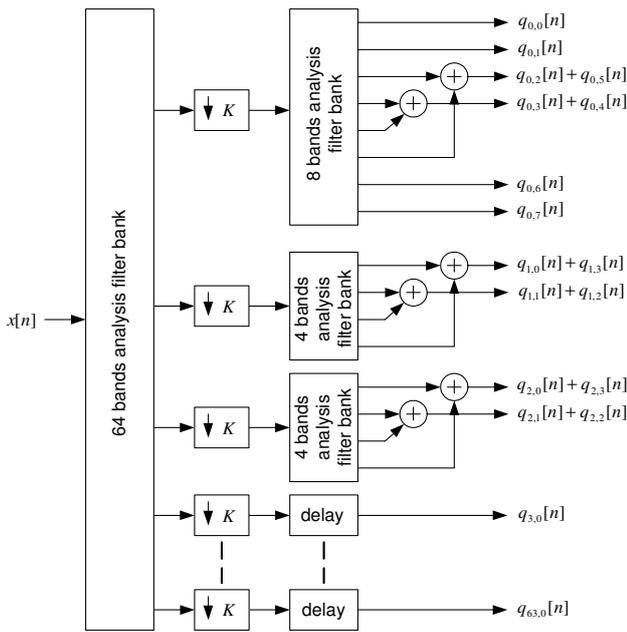
Figure 4: *Block diagram of hybrid analysis filter bank with $M_0 = 8$, $M_1 = 4$ and $M_2 = 4$, some of the filter bank outputs have been combined to reduce complexity.*

are obtained according to Equation 4. Finally, two instances of the hybrid synthesis filter bank are used to convert the reconstructed stereo signal into the time domain.

The stereo parameters are typically updated every 16 or 32 QMF samples, i.e., every 21.3 or 42.7 ms at 48 kHz sampling rate. In order to achieve smooth transitions, linear interpolation is applied to the elements $h_{i,j}$ of the matrix **H** in Equation 4 for the QMF samples located between two stereo parameter updates.

Complexity estimates for the earlier FFT-based PS decoder and low complexity QMF-based PS decoder indicate that the complexity of the FFT-based system is dominated by the t/f and f/t transforms. The QMF-based decoder for 20 stereo bands reduced the computational complexity by 42% and the RAM requirements by 80% when compared to the FFT-based PS decoder.

4. GENERATION OF SYNTHETIC AMBIENCE

There are various ways of generating a decorrelated signal d from a mono signal m to enable the stereo reconstruction outlined in Section 2.2. The common approach is to apply an all-pass filter to m. An obvious such all-pass filter is a constant delay, and typically a delay time like 10 ms is used. However, when m and d are combined in the mixing process of Equation 4, the reconstructed signals $\hat{l}, \hat{r}$ can exhibit a strong comb-filter like characteristic.

To reduce this problem, a frequency dependent delay can be used, which corresponds to a filter with a chirp-like impulse response [15]. A better and more advanced approach to the decorrelation problem utilizes principles known from artificial reverberation systems and will be presented below. A more detailed discussion of synthetic ambience in parametric stereo coding can be found in [4].

4.1. The QMF/IIR all-pass approach

To achieve computational efficient artificial reverberation of an audio signal, algorithms combining delay lines of different lengths with feedback and all-pass filters are commonly used. Such an algorithm, configured for a very short reverberation time of a few 10 ms, can also be seen as an IIR all-pass filter and constitutes a powerful decorrelation filter.

When used in combination with QMF-based PS tool discussed in this paper, it is advantageous to implement the decorrelation for the QMF sub-band signals. This allows to easily change the decorrelator characteristics over frequency. To reduce the computational complexity, it is also possible to use a simple frequency depended delay for QMF sub-bands at high frequencies. In the system discussed here, a reverb-based decorrelation is employed only for the first 23 QMF bands.

Choosing delay line lengths is a crucial parts in reverberator design. Best results are usually obtained by using lengths that are large numbers and are mutually prime. This is a problem in the QMF domain because of the low sampling rate of the QMF sub-band signals (750 Hz for a 48 kHz original signal). To access finer time resolution, delay by fractions of the sampling interval was employed. Such a fractional delay can easily be approximated by rotating the phase of the complex (approximately analytic) QMF sub-band signals by the angle that corresponds to the desired fraction of a sample at the center frequency of the QMF band in question.

The resulting QMF/IIR decorrelator consists of a chain of three delay lines with the length of 3, 4, and 5 QMF samples. Each delay line includes also a fractional delay and the feedback is implemented in an all-pass like manner with appropriately chosen filter coefficients. Figure 5 (b) indicates the impulse response of the complete QMF/IIR decorrelator in the time domain, i.e., after the QMF synthesis bank.

4.2. Improving transient behavior

A major problem when incorporating delays or all-pass filters that include long decays into a decorrelator is the performance at transients due to the risk of generating audible post-echoes or unnatural colorization. An efficient approach to overcome this problem is to detect such transients in the decoder, i.e., within the PS synthesis, and thereafter reduce the level of the decorrelated signal using a soft decision.

Figure 5 shows the behavior of the transient reducing process. Waveform (a) shows the original input signal m used in this experiment. It is a short impulse that is followed by a long tail of reverb (room response) present in the original downmixed signal. Waveforms (b) and (c) show the sum $m + d$ of the original and the decorrelated signals (i.e., synthetic ambience), without and with the transient reduction, respectively. It can be seen that the impulse response of the decorrelator is strongly attenuated when applying the transient reduction process. Hence, the reverberation characteristics (e.g. perceived room size) of the original signal remains practically unchanged.

5. COMBINING PS WITH AACPLUS IN MPEG-4

The combination of MPEG-2/4 AAC with the SBR bandwidth extension tool is known as aacPlus and was standardized in MPEG-4 as HE-AAC [8]. The basic principle of SBR has been elaborated
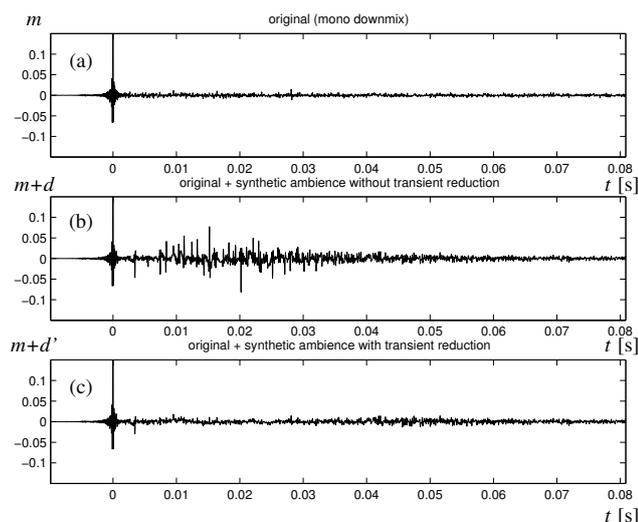
Figure 5: *Impulse fed to a conventional synthetic reverberator. (a) Original mono downmixed signal (no decorrelation). (b) Original signal + QMF/IIR decorrelation without transient reduction. (c) Original signal + QMF/IIR decorrelation with transient reduction.*

Figure 6: *(a) Block diagram of an aacPlus v1 decoder. (b) Block diagram of an aacPlus v2 decoder, i.e., aacPlus with parametric stereo.*

on in several papers [6], [16], [14]. Figure 6 (a) depicts a mono aacPlus decoder. The output of the AAC decoder, which operates at half the sampling rate of the full bandwidth audio signal, is rst analyzed with a 32 band QMF bank. Then a HF generator recreates the highband by patching QMF sub-bands from the lowband to the highband. An envelope adjuster modi es the spectral envelope of the regenerated highband and can add additional components such as noise and sinusoids, all according to the SBR guidance information in the bit stream. Finally the lowband and highband are combined and a 64 band QMF synthesis bank is employed to obtain full bandwidth decoder output signal at the original sampling rate.

5.1. Overview of the complete AAC+SBR+PS system

When the PS tool presented in this paper is combined with aacPlus, this results in a coder that achieves a signi cantly increased coding ef cienc y for stereo signals at very low bit rates when compared to aacPlus operating in normal stereo mode. Figure 6 (b) shows a simpli ed block diagram of the resulting decoder, which is referred to as aacPlus v2. Since the SBR tool already operates in the QMF domain, the PS tool can be included in such a decoder in a computationally very ef cient manner directly prior to the nal QMF synthesis lter bank. Comparing Figures 6 (a) and (b), it is evident that only the parametric stereo decoding and synthesis, including its hybrid lter bank, have to be added to a mono aacPlus decoder, plus of course a second QMF synthesis 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 ltering and SBR processing have to be carried out for both channels of a stereo signal.

The PS tool allows for e xible con guration 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
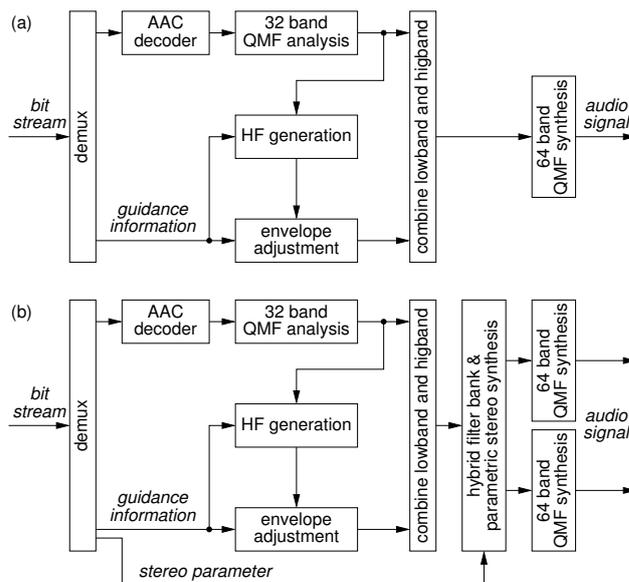
codebooks, makes it possible to operate this PS system over a large range of bit rates.

When an aacPlus v2 coder is operated at target bit rate of 24 kbit/s, the PS parameters require an average side information bit rate of 2 to 2.5 kbit/s, assuming 20 stereo bands for IID and ICC. For lower target bit rates, the PS frequency resolution can be decreased to 10 bands, reducing the PS side information bit rate accordingly. On the other hand, the PS tool permits to increase time and frequency resolution and to transmit IPD/OPD parameters, which improves the quality of the stereo reconstruction at the cost of 10 kbit/s or more PS side information.

5.2. Subjective test results

Figure 7 shows subjective results from a listening test comparing aacPlus v1 using normal stereo coding at 24 and 32 kbit/s with aacPlus v2 utilizing the PS tool at 24 kbit/s [13]. 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 aacPlus stereo veri cation test [17] were used as test material and playback was done using headphones. The test employed MUSHRA [18] methodology and included a hidden reference and low-pass ltered anchors with 3.5 and 7 kHz bandwidth.

At both test sites, it was found that aacPlus v2 at 24 kbit/s achieves an average subjective quality that is equal to aacPlus v1 stereo at 32 kbit/s and that is signi cantly better than aacPlus v1 stereo at 24 kbit/s. It is of interest to relate these results to the MPEG-4 veri cation test [17]. There, it was found that aacPlus v1 stereo at 32 kbit/s achieved a subjective quality that was signi cantly 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 aacPlus v2 achieves more than twice the coding ef cienc y of AAC for stereo signals. Further MUSHRA tests have shown that aacPlus v2 achieves a signi cantly better subjective quality than aacPlus v1 stereo also for 18 and 32 kbit/s.
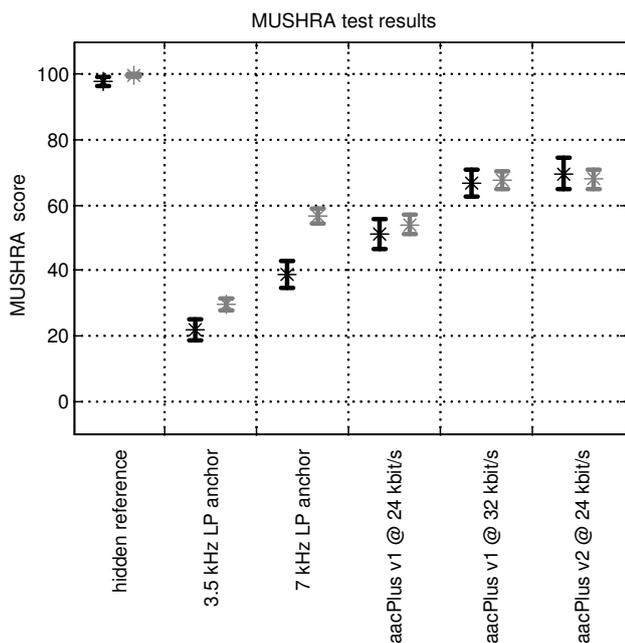
Figure 7: *MUSHRA listening test results for two sites (black and gray) showing mean grading and 95% con dence interval (from [13]).*

6. CONCLUSIONS

A low complexity parametric stereo coding tool has been presented. It was shown that this parametric stereo coding tool signi cantly enhances the coding ef cienc y of existing audio coders. The presented tool is particularly interesting in combination with audio coders using SBR bandwidth extension, since the resulting coder has approximately the same computational complexity as in a normal stereo con guration. The combination of AAC, SBR, and the parametric stereo tool presented here was included in the MPEG-4 Audio standard and is referred to as aacPlus v2. It enables coding of stereo signals at bit rates that are less than 50% of those required by AAC to achieve the same subjective quality and was recently adopted as recommended coder for audio streaming in Release 6 of the 3GPP standard for mobile services.

7. ACKNOWLEDGMENTS

The author would like to express his gratitude to Erik Schuijers, Jeroen Breebaart, and Jonas Engdegard for providing some of the material that was used for preparing this paper.

8. REFERENCES

[1] J. Herre and H. Purnhagen, "General audio coding," in *The MPEG-4 Book*, F. Pereira and T. Ebrahimi, Eds., chapter 11. Prentice Hall, Englewood Cliffs, NJ, US, 2002.

[2] C. Faller and F. Baumgarte, "Binaural cue coding: A novel and ef cient representation of spatial audio," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Orlando, FL, US, May 2002.

[3] E. Schuijers, W. Oomen, B. den Brinker, and J. Breebaart, "Advances in parametric coding for high-quality audio," in *AES 114th Convention*, Amsterdam, NL, Mar. 2003, Preprint 5852.

[4] J. Engdegard, H. Purnhagen, J. Rödén, and L. Liljeryd, "Synthetic ambience in parametric stereo coding," in *AES 116th Convention*, Berlin, DE, May 2004.

[5] 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.

[6] M. Dietz, L. Liljeryd, K. Kjörling, and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," in *AES 112th Convention*, Munich, DE, May 2002, Preprint 5553.

[7] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 Advanced Audio Coding," *J. Audio Eng. Soc.*, vol. 45, no. 10, pp. 789–814, Oct. 1997.

[8] 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.

[9] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "High-quality parametric spatial audio coding at low bitrates," in *AES 116th Convention*, Berlin, DE, May 2004.

[10] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegard, "Low complexity parametric stereo coding," in *AES 116th Convention*, Berlin, DE, May 2004.

[11] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," *EURASIP J. on Applied Signal Processing*, 2004, (submitted).

[12] W. Oomen, E. Schuijers, H. Purnhagen, and J. Engdegard, "MPEG4-Ext2: CE on low complexity parametric stereo," ISO/IEC JTC1/SC29/WG11 MPEG2003/M10366, Dec. 2003.

[13] H. Purnhagen, J. Engdegard, W. Oomen, and E. Schuijers, "Combining low complexity parametric stereo with High Ef cienc y AAC," ISO/IEC JTC1/SC29/WG11 MPEG2003/M10385, Dec. 2003.

[14] 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.

[15] 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.

[16] M. Wolters, K. Kjörling, D. Homm, and H. Purnhagen, "A closer look into MPEG-4 High Ef cienc y AAC," in *AES 115th Convention*, Los Angeles, USA, Oct. 2003, Preprint 5871.

[17] ISO/IEC JTC1/SC29/WG11, "Report on the veri cation tests of MPEG-4 High Ef cienc y AAC," ISO/IEC JTC1/SC29/WG11 N6009, Oct. 2003.

[18] ITU-R, "Method for the subjective assessment of intermediate quality level of coding systems (MUSHRA)," ITU-R Recommend. BS.1534, 2001.