Hindawi Publishing Corporation
EURASIP Journal on Audio, Speech, and Music Processing
Volume 2009, Article ID 468971, 21 pages
doi:10.1155/2009/468971

Review Article

An Overview of the Coding Standard MPEG-4 Audio Amendments 1 and 2: HE-AAC, SSC, and HE-AAC v2

A. C. den Brinker,[1] J. Breebaart,[1] P. Ekstrand,[2] J. Engdegård,[2] F. Henn,[2] K. Kjörling,[2] W. Oomen,[3] and H. Purnhagen[2]

[1] *Philips Research Laboratories, High Tech Campus 36, 5656 AE Eindhoven, The Netherlands*
[2] *Dolby Sweden AB, Gävlegatan 12 A, 11330 Stockholm, Sweden*
[3] *Philips Applied Technologies Eindhoven, High Tech Campus 5, 5656 AE Eindhoven, The Netherlands*

Correspondence should be addressed to A. C. den Brinker, bert.den.brinker@philips.com

Received 29 September 2008; Accepted 24 February 2009

Recommended by James Kates

In 2003 and 2004, the ISO/IEC MPEG standardization committee added two amendments to their MPEG-4 audio coding standard. These amendments concern parametric coding techniques and encompass Spectral Band Replication (SBR), Sinusoidal Coding (SSC), and Parametric Stereo (PS). In this paper, we will give an overview of the basic ideas behind these techniques and references to more detailed information. Furthermore, the results of listening tests as performed during the final stages of the MPEG-4 standardization process are presented in order to illustrate the performance of these techniques.

1. Introduction

The MPEG-2 Audio coding standard was released in 1997 and has successfully found its way into the market. Later, MPEG-4 Audio Version 1 and Version 2 were issued in mid 1999 and early 2000, respectively. These versions have adopted the MPEG-2 AAC coder including several extensions to it. In addition several other components like speech coders (HVXC and CELP) and a Text-To-Speech Interface are specified.

In 2001, MPEG identified two areas for improved audio coding technology and issued a Call for Proposals (CfP, [1]). These two areas were

(i) improved compression efficiency of audio signals or speech signals by means of bandwidth extension which is forward and backward compatible with existing MPEG-4 technology;

(ii) improved compression efficiency of high-quality audio signals by means of parametric coding.

This started a new cycle in the standardization process, consisting of a competitive phase leading to a selection of the reference model, a collaborative phase for improving the technology and, finally, the definition of a new standard. Close to the finalization of the work on parametric coding, it was demonstrated that the parametric stereo (PS) module that was developed in the course of the this work item could also be combined with the bandwidth extension technology thereby providing a significant additional boost in coding efficiency. This particular combination was subsequently added to the parametric coding amendment. The work on bandwidth extension and parametric coding reached the final stage of Amendment 1 and 2 to MPEG-4 Audio mid 2003 and 2004, respectively.

This paper has the intention to outline the ingredients of the MPEG-4 Audio Amendments in a comprehensive way. Part of this material is present in the literature but mostly scattered. Therefore, this paper sets out to give an overview of the three components that make up these Amendments with references to more detailed information where necessary.

The outline of the paper is as follows. The basic technology and subjective test results for the bandwidth extension and parametric coding are discussed in Sections 2 and 3, respectively. In Section 4, the combination of AAC, SBR and PS is outlined, including subjective test results. Finally, in Section 5, the conclusions are presented.

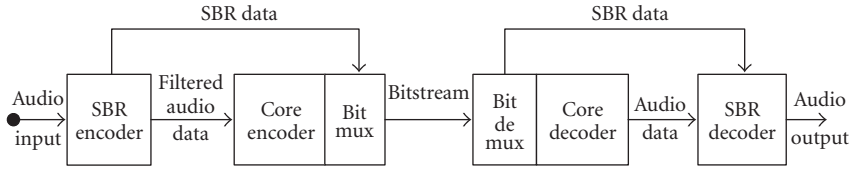

Figure 1: SBR incorporates pre- and postprocessing stages: SBR parameters are extracted from the input signal in the SBR encoder and transmitted together with the encoded low-band signal. The decoded low-band signal is bandwidth extended by the SBR decoder, spectrally adjusted and enhanced in accordance with the transmitted SBR data.

2. MPEG-4 SBR

2.1. Introduction. High-Frequency Reconstruction/Regeneration (HFR), or BandWidth Extension (BWE), techniques have been researched in the speech coding community for decades [2, 3]. The underlying hypothesis stipulates that it should be possible to reconstruct the higher frequencies of a signal given the corresponding low-frequency content only. In the speech coding community this research was done with the goal to be able to accurately reconstruct the high-band of a speech signal given only the low-pass filtered low-band signal and no other a priori information about the high-band of the original signal. Typically the high-band was recreated by upsampling of the low-band signal without subsequent low-pass filtering (aliasing), or by means of broad-band frequency translation (single side-band modulation) of the low-band signal [2, 3]. The spectral envelope of the recreated high-band was either simply whitened and tilted with a suitable roll-off at higher frequencies, or in more elaborate versions [4] estimated by means of statistical models. This research has not led to any wide adoption of such an HFR-based speech enhancement in the market as of today.

The original SBR technique (of which the development started in early 1997) differs from previously known HFR techniques [5, 6].

(i) The primary means for extending the bandwidth is transposition, which ensures that the correct harmonic structure is maintained for single- and multipitched signals alike.

(ii) Spectral envelope information is always sent from the encoder to the decoder making sure that the spectral envelope of the reconstructed high-band is correct [7].

(iii) Additional means such as inverse filtering, noise, and sinusoidal addition, guided by transmitted information, compensate for shortcomings of any bandwidth extension method originating from occasional fundamental dissimilarities between low-band and high-band [8, 9].

These features successfully enabled the use of a bandwidth extension technique not only for speech signals but for arbitrary signals. The fundamental topology of a system employing SBR is shown in Figure 1. An audio input signal is first processed by an SBR encoder, resulting in a low-pass filtered audio signal and SBR data. The audio signal is subsequently encoded using a core encoder. Finally, the SBR data and the core-coder output are combined into an output bit stream. The decoder performs the reverse process.

Since the HFR method enables a reduction of the core coder bandwidth and the HFR technique requires significantly lower bit rate to code the high-frequency range than a waveform coder would, a coding gain can be achieved by reducing the bit rate allocated to the waveform core coder while maintaining full audio bandwidth. Naturally, this gives the possibility to decrease the total data rate by lowering the crossover frequency between core coder and the HFR part. However, since the audio quality of the HFR part cannot scale towards transparency, this crossover frequency is always a delicate tradeoff between core coder and HFR related artifacts.

This paper only covers SBR in the MPEG context, where it is standardized for use together with AAC, forming the (High Efficiency) HE AAC Profile. However, the algorithm and bit stream are essentially core codec agnostic, and SBR has successfully been applied to other codecs such as MPEG Layer-2 [10] and MPEG Layer-3 (the latter case is known as mp3PRO, see [11]), it is included in (High Definition Codec) HDC, that is, the proprietary codec used by iBiquity, and is standardized within (Digital Radio Mondiale) DRM for use together with the CELP and HVXC speech codecs [12]. Furthermore, it is worth noting that the transposition method included in the MPEG-4 standard is a carefully selected tradeoff between implementation cost and quality, relaxing the strict requirements on harmonic continuation that are met by more advanced transposition methods.

2.2. System Overview

2.2.1. SBR Encoding Process

Overview. Following the general process of MPEG to standardize transmission formats and decoder operation (and hence allowing future encoder-side improvements) the SBR amendment contains an informative (as opposed to normative) encoder description. Hence this section gives a generic overview of the various elements of an encoder; the exact design of these elements is left up to the implementer. However, for detailed information on a realization of the encoder capable of high perceptual performance, the 3GPP specification of the SBR encoder is a good source, see [13].

The basic layout of an SBR encoder is depicted in the block diagram of Figure 2. Central to the operation of both encoder and decoder are dedicated, complex-valued

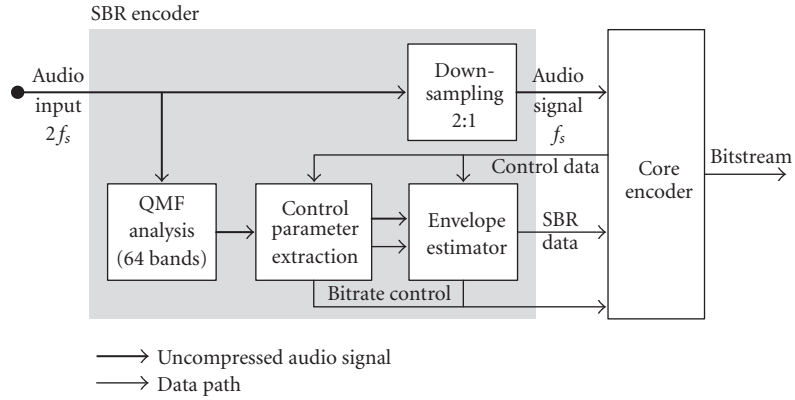

Figure 2: Block scheme of the SBR encoder. The core of the system comprises a QMF bank and a spectral envelope estimator.

filter banks of the (Quadrature Mirror Filter) QMF type. The encoder has an analysis bank per input channel, and the decoder has an analysis and synthesis pair per channel. Most of the SBR processing, such as encoder-side parameter extraction and decoder-side bandwidth extension and spectral envelope adjustment, is performed in the QMF domain.

QMF Analysis. The original time-domain input signal is first filtered in a 64-channel analysis QMF bank. The filter bank splits the time-domain signal into complex-valued subband signals and is thus oversampled by a factor of two compared to a regular real-valued QMF bank [14]. For every 64 time-domain input samples, the filter bank produces 64 subband samples. At 44.1 kHz sample rate this corresponds to a nominal bandwidth of 344 Hz, and a time resolution of 1.4 ms. All the subsequent modules in the encoder operate on the complex-valued subband samples.

Transient Detection. A transient detector (part of the "Control parameter extraction" in Figure 2) operates on the complex-valued subband signals in order to assist the envelope estimator in the time/frequency (T/F) grid selection. Generally, longer time segments of higher frequency resolution are produced by the envelope estimator during quasistationary passages, while shorter time segments of lower frequency resolution are used for dynamic passages. The transient detection is, for example, accomplished by calculating running short-term energies and detecting significant changes.

T/F Grid Selection and Envelope Estimation. The estimated envelope data are obtained by averaging of subband sample energies within segments in time and frequency. The time borders of these segments are determined mainly by the output from the transient detector, and are subsequently signaled to the decoder. When the transient detector signals a transient to the envelope estimator, segments of shorter duration in time are defined by the envelope estimator, starting with a minimal segment, the leading border of which is placed at the onset of the transient. Subsequent to the short-time segment by the transient, somewhat longer segments are used to correctly track a potential decay of the transient, and finally long segments are used for the stationary part of the signal.

The main objective is to avoid pre- and postechoes that otherwise would be induced by the envelope adjustment process in the decoder for transient input signals.

The envelope estimator also decides on the frequency resolution to use within each time segment. The variable frequency resolution is achieved by employing two different schemes for grouping of QMF samples in frequency: high resolution and low resolution, where the number of estimates differs by a factor of two. In order to reduce instantaneous peaks in the SBR bit rate, the envelope estimator typically trades one high-resolution envelope for two low resolution ones. The grouping in frequency can be either linearly spaced or (approximately) log spaced where the number of bands to use per octave is variable. An example of a T/F grid selection is given in Figure 3 where the grid is superimposed on a spectrogram of the input signal. As is clear from the figure, the time resolution is higher around the transient events, albeit with lower frequency resolution, and vice versa for the more stationary parts of the signal.

Although the segment borders can be chosen with a high degree of freedom, the temporal resolution, as well as the frequency resolution, is constrained by the analysis QMF bank resolution. The filter bank is designed to provide a resolution in both time and frequency that is considered adequate for the adjustment of the envelope for all signal types. Hence the filter bank resolution is not adaptive, as is usually the case for filter banks in perceptual waveform coders, and the estimates are achieved by, within a filter bank of fixed size, adaptively grouping and averaging of subband sample energies as outlined above.

Noise Floor Estimation. An important aspect of the SBR encoder is to assess to which extent the tonal-to-noise ratio of the reconstructed high-band will be correct. For this purpose, the encoder estimates the amount of additional noise that needs to be added at the decoder side after

regeneration of the high-band. This is done in an analysis-by-synthesis fashion. In Figure 4 such an analysis-by-synthesis process is illustrated. In the top panel of the figure a spectrum of the input signal is given. In this particular example the input signal is a synthetically generated test signal of which the tonal (harmonic) structure ends abruptly above 5.5 kHz. The remaining spectrum of the signal consists of noise. In the lower panel of the figure a spectrum is given of the high-band given the HF generation method used in the decoder, *without* additional correction of tonal-to-noise properties. In this case, the tonal structure of the low-band has propagated to the high-band (the region from 5.5 kHz to 15 kHz) and hence within the region of 5.5 to 15 kHz, there is a mismatch in signal characteristics between original input and reconstructed high-band signal. The transmission of additional noise information allows correction of such mismatches. It should be noted that the spectrum in the lower panel illustrates the low-band signal in combination with the high-band signal after HF generation without any subsequent envelope adjustment.

Missing Harmonics Detection. Similarly to the above situation, the encoder also needs to assess whether strong tonal components in the original high-band signal will be missing after the high-frequency reconstruction. In Figure 5 an example is given where three strong tonal components are not reconstructed by the high-frequency regeneration based on the low-band signal. Again an analysis-by-synthesis approach can be beneficial. For this example a glockenspiel signal is used. In the upper panel of Figure 5 the spectrum for the input signal is given, where three strong tonal components in the high-band are indicated by circles. In the lower panel of Figure 5 the spectrum of the HF-generated signal is given similarly to the example in Figure 4. Clearly the three strong tonal components will not be properly regenerated by the HF generator, and therefore need to be replaced by sinusoids generated separately in the decoder. Information on the (frequency) location of these strong tonal components is transmitted to the decoder, and the missing components are inserted in the high-band signal.

Quantization and Encoding. The SBR envelope data, tonal component data, and noise-floor data are quantized and differentially coded in either the time or frequency direction in order to minimize the bit rate. All data is entropy coded using Huffman tables. Details about SBR data coding are given in the next section.

2.2.2. SBR Bit Stream

Overview. To ensure consistent coding of transients regardless of localization within codec frames, the SBR frames have variable time boundaries, that is, the exact duration in time covered by one SBR frame may vary from frame to frame. The bit stream is designed for maximum flexibility such that it scales well from the lowest bit rate applications up to medium and high bit rate use cases, and is easy to adapt for different core codec frame lengths. Furthermore, it is possible to trade bit-error robustness against increased coding efficiency by selecting the degree of interframe dependencies, and the signaling scheme offers error detection capabilities in addition to a Cyclic Redundancy Check (CRC).

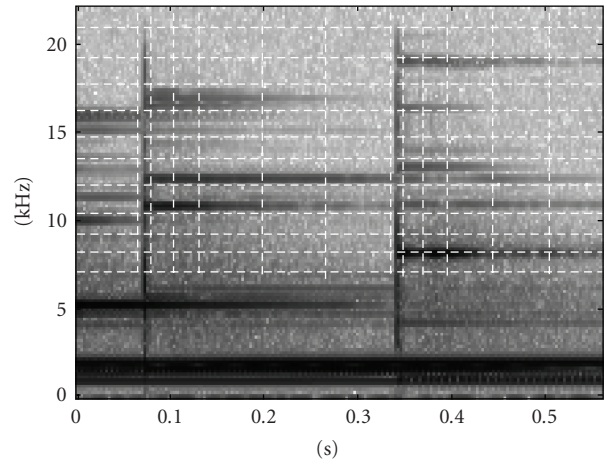

Figure 3: T/F grid selection example. The white dashed lines illustrate the borders of the time-frequency tiles superimposed on the spectrogram of the input signal. The leading edge and decay of the transient is encoded with short low frequency resolution envelopes, and the quasistationary passages in between transients are represented by longer high-frequency resolution envelopes.

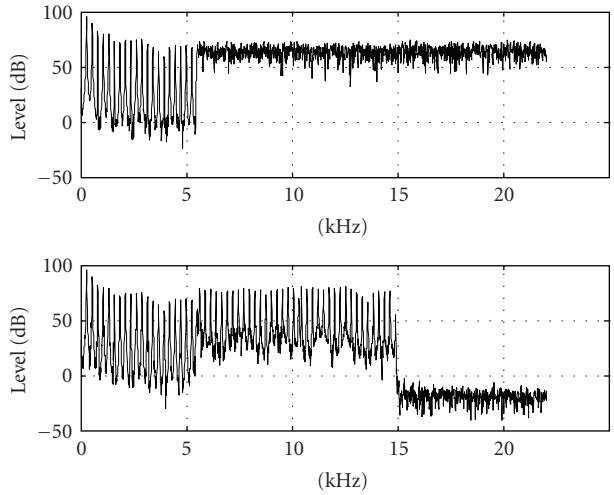

Figure 4: Illustration of the mismatch in noise level of the reconstructed high-band if no additional noise information is transmitted to the decoder. This can be used for an analysis-by-synthesis method in the SBR encoder in order to assess the amount of noise that should be added on the decoder side. The upper panel shows the spectrum of the (synthetically generated) input signal, and the lower panel shows a spectrum of the signal obtained after HF generation based on the low-band signal without noise correction. The SBR range covers the frequency range from 5.5 kHz to 15 kHz.

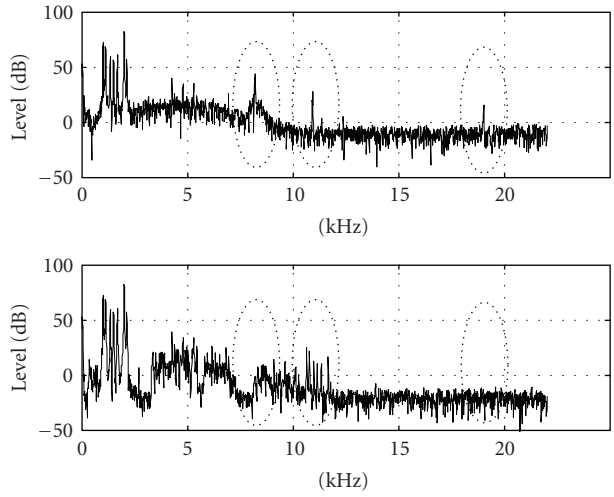

Figure 5: Illustration of missing sinusoidal components in the high-band (if no additional sinusoidal signals are added in the decoder). This can be used for an analysis-by-synthesis method in the SBR encoder in order to assess where a sinusoid should be added on the decoder side. The upper panel shows the spectrum of the input signal (a glockenspiel signal), and the lower panel shows a spectrum of the signal obtained after HF generation based on the low-band signal without adding separate sinusoids. The SBR range covers the frequency range from 5.5 kHz to 15 kHz.

2.2.3. SBR Decoding Process

Overview. The block scheme of the SBR decoder is given in Figure 6. The bit stream is input to the core decoder providing the low-band signal, and the SBR relevant bit stream to the SBR decoder. The SBR decoder performs a 32 subband analysis of the low-band signal, which is subsequently used, along with control data from the bit stream, by the HF generator to create the high-band signal. The envelope of the recreated high-band signal is subsequently adjusted and additional signal components are added to the high-band. The combined low-band and high-band are finally synthesized by a 64 subband QMF synthesis filter bank in order to obtain the time-domain output signal. The analysis and synthesis filter banks are constructed such that an upsampling of the low-band signal by a factor of two is inherently obtained in the processing. A detailed description of the decoder can be found in the MPEG-4 Audio standard [15]. In the following, we merely outline the various decoding steps.

An example is given in Figure 7. The original input signal spectrum is shown in the top-left panel. The spectrum of a low-band output from the AAC core decoder is given in the top right panel of Figure 7. It is clear that the signal is low-pass filtered at approximately 6 kHz which is the bandwidth covered by the core coder for the setting corresponding to the bit rate used in this example. It should be noted that in the figure the signal has been upsampled to the sampling frequency of the original signal (and also that of the final output signal) in order to allow for spectrum comparison.

The HF Generator transposes parts of the low-band frequency range to the high-band frequency range covered by SBR as indicated in the bit stream. In the bottom left panel of Figure 7 the spectrum of the transposed intermediate signal in combination with the low-band signal is displayed. This is how the output would look if no envelope adjustment of the recreated high-band would be performed.

The envelope adjuster adjusts the spectral envelope of the recreated high-band signal according to the envelope data and time/frequency grid that was transmitted in the bit stream. Additionally, noise and sinusoid components are added as signaled in the bit stream. The output from the SBR decoder after envelope adjustment is depicted in the bottom right panel of Figure 7. In the following the decoding steps are examined in more detail.

QMF Analysis. The time-domain audio signal, supplied by the core decoder and usually sampled at half the frequency of the original signal, is first filtered in the analysis QMF bank. The filter bank splits the time-domain signal into 32 subband signals. For every 32 time-domain samples, the filter bank produces 32 complex-valued subband samples and is thus over-sampled by a factor of two compared to a regular real-valued QMF bank. The oversampling enables significant reduction of impairments emerging from modifications of subband signals. The oversampling is accomplished through extension of a cosine modulated filter bank with an imaginary sine modulated part, forming a complex-exponential modulated filter bank. In a conventional cosine modulated filter bank the analysis and synthesis filters $h_k(n)$ and $f_k(n)$ are cosine modulated versions of a symmetric low-pass prototype filter $p_0(n)$ as

$$\begin{aligned} h_k(n) &= 2p_0(n)\cos\left\{\frac{\pi}{2M}(2k+1)\left(n-\frac{N}{2}-\frac{M}{2}\right)\right\}, \\ f_k(n) &= 2p_0(n)\cos\left\{\frac{\pi}{2M}(2k+1)\left(n-\frac{N}{2}+\frac{M}{2}\right)\right\}, \end{aligned} \tag{1}$$

where $k = 0 \cdots M-1$, M is the number of channels and $n = 0 \cdots N$, where N is the prototype filter order. Figure 8 depicts a simplified block scheme for the implementation of a cosine modulated filter bank. For complex modulation both filters are obtained from

$$h_k(n) = f_k(n) = p_0(n)\exp\left\{\frac{\pi}{2M}(2k+1)\left(n-\frac{N}{2}\right)\right\}. \tag{2}$$

The terms containing $M/2$ (terms needed for aliasing cancellation) present in the traditional cosine modulated filter bank omitted because of the complex-valued representation [14]. In Figure 9 the corresponding block scheme for a complex-valued filter bank implementation is outlined. The complex-exponential modulation creates complex-valued subband signals that can be interpreted as the analytic versions of the signals obtained from the real part of the filter bank. This feature provides a subband representation suitable for various modifications, and also an inherent measure of the instantaneous energy for the subband signals [14]. The prototype filter used for HE-AAC is of order 640 (N) and gives a reconstruction error of −65 dB.

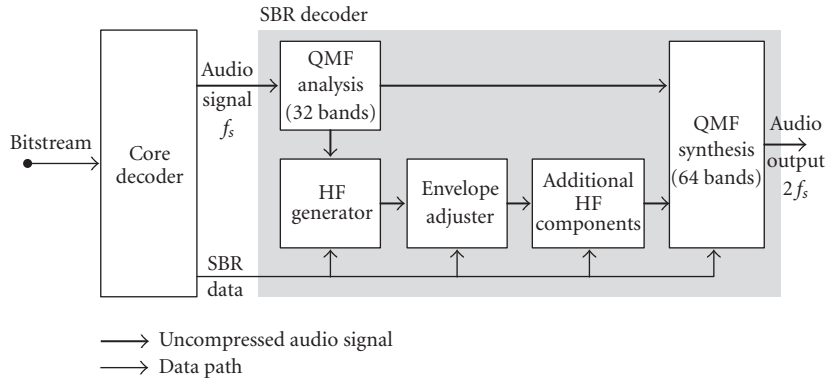

Figure 6: Block scheme of the SBR decoder. The received bit stream is input to the core decoder decoding the low-band audio signal, and providing the SBR decoder with the SBR relevant bit stream data. The SBR decoder performs a QMF analysis of the low-band signal which is subsequently used for the HF Generation providing a high-band signal. The high-band is envelope adjusted and additional signal components are added. Finally, the output signal is obtained by a QMF synthesis filter bank.

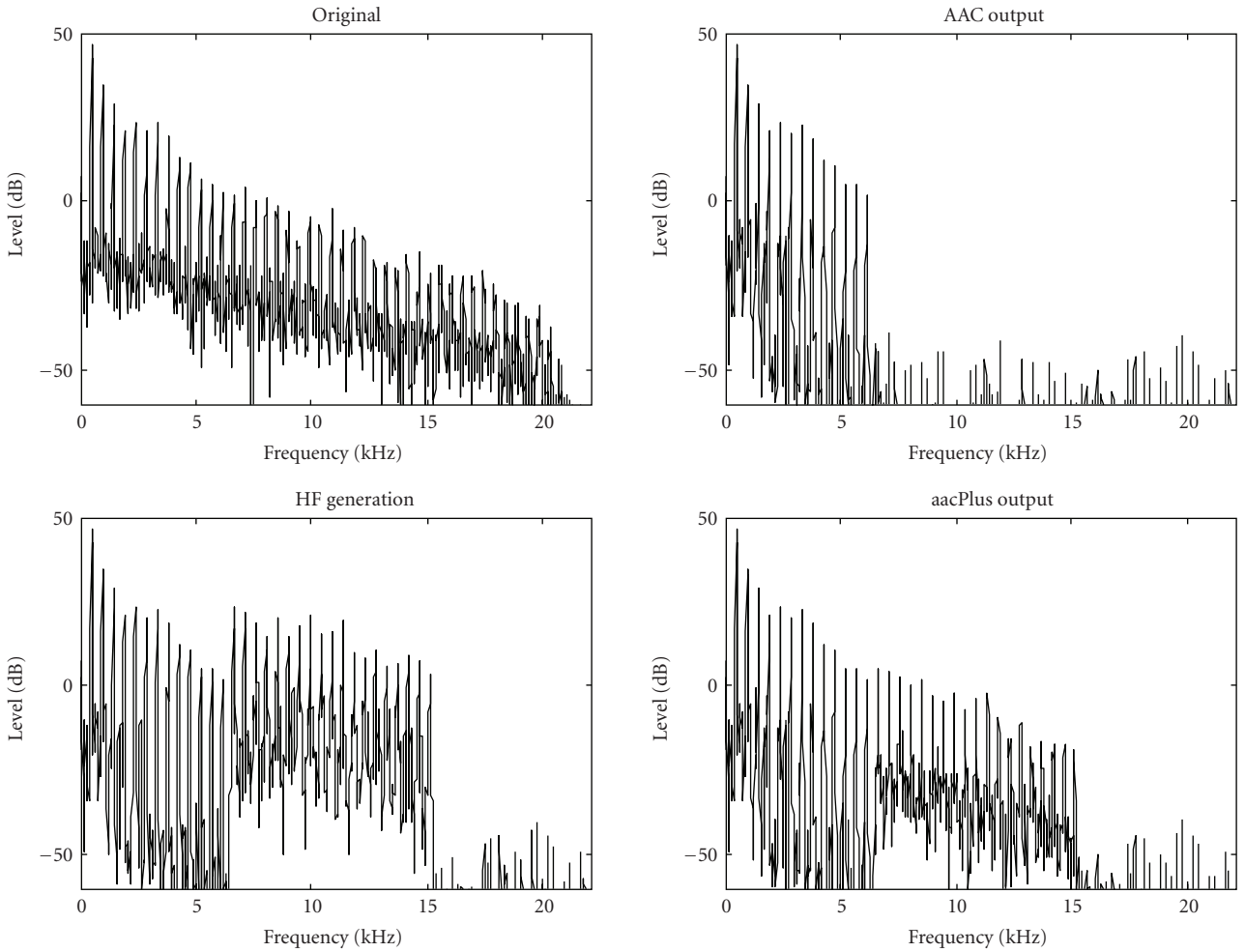

Figure 7: Spectrum of the signal at different points of processing in the SBR decoder. Top left is the (power) spectrum of the original signal, top right is the spectrum of the low-band signal resulting from the AAC decoder, bottom left is the spectrum of the combined low-band and high-band prior to envelope adjustment, bottom right is the spectrum of the output signal.

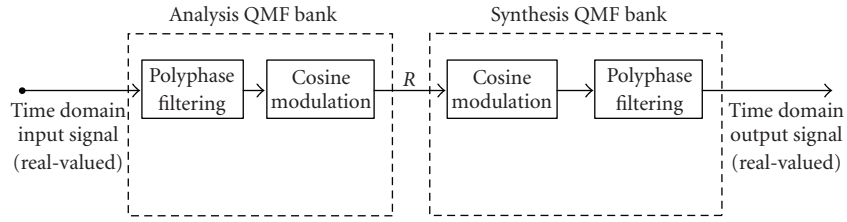

Figure 8: Simplified block scheme of a Cosine Modulated QMF bank implementation.

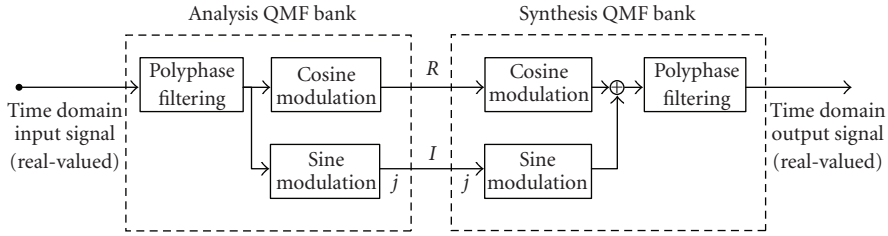

Figure 9: Simplified block scheme of a Complex Exponential Modulated QMF bank implementation.

HF Generation. The complex-valued subband signals obtained from the filter bank are processed in the high-frequency generation unit to obtain a set of high-band subband signals. The generation is performed by selecting low-band subband signals, according to specific rules, which are mirrored or copied to the high-band subband channels. The patches of QMF subband to be copied, their source range and target range, are derived from information on the borders of the SBR range, as indicated by the bit stream. The algorithm generating the patch structure has the following objectives.

(i) The patches should cover the frequency range up to 16 kHz with as few patches as possible, without using the QMF subband lowest in frequency (i.e., the subband including DC) in any patch.

(ii) If several patches constitute the high-band, a patch covering a lower frequency range should have a wider or equal bandwidth compared to a patch covering a higher frequency range. The motivation is that for lower frequencies the human hearing is more sensitive, and therefore patches with wide bandwidth are preferred for lower frequencies in order to move any potential discontinuity between the first and the second patch as high up in frequency as possible.

(iii) The source frequency range for the patches should be as high up in frequency as possible.

Creating the high-band in this way has several advantages and is the reason why SBR can be referred to as a semi-, or quasi-, parametric method. Although the high-band is synthetically generated and shaped by the SBR bit-stream data, the characteristics of the high-band are inherited from the low-band, and, which is the most important aspect, so is the temporal structure of the high-band. This makes the corrections of the high-band, in order to resemble the original, much more likely to succeed in the subsequent processing steps.

With the above in mind, the characteristics of the low-band and the high-band still vary for different audio signals. For example, the tonality is usually more pronounced in the low-band than in the high-band. Therefore, inverse filtering is applied to the generated high-band subband signals. The filtering is accomplished by in-band filtering of the complex-valued signals using adaptive low-order complex-valued FIR filters. The filter coefficients are determined through an analysis of the low-band in combination with control signals extracted from the SBR data stream. A second-order linear predictor is used to estimate the spectral whitening filter using the covariance method. The amount of inverse filtering is controlled by a chirp-factor given from the bit stream. Hence, the HF-generated signal $y_k(n)$ for QMF subband k and time slot n in the high-band can be defined according to

$$y_k(n) = x_l(n) - \alpha_l(0)\gamma_k x_l(n-1) - \alpha_l(1)\gamma_k^2 x_l(n-2), \quad (3)$$

where $\alpha_l(0)$ and $\alpha_l(1)$ are given by the prediction error filter estimated for the low-band subband l, and where γ_k is the chirp factor (between 0 and 1) controlled by the bit stream.

In Figure 10 an example of patching and inverse filtering is given. In the top panel of the figure, a (power) spectrum of the low-band signal is displayed, and the maximum source region for the patching is indicated. For all subbands within this region, prediction error filters are estimated as outlined above. The source range in the low-band is patched, in this example, to region A and B. The frequency plot of the patched signals in these regions are given in the lower panel of Figure 10. Here three inverse filtering regions are also indicated by 1, 2, and 3. The applied inverse filtering level is the same within these regions and its parameters are contained in the bit stream.

Given that the subband signals are patched from the low-frequency region to region A and B in Figure 10, so are the prediction error filter coefficients for the low-frequency

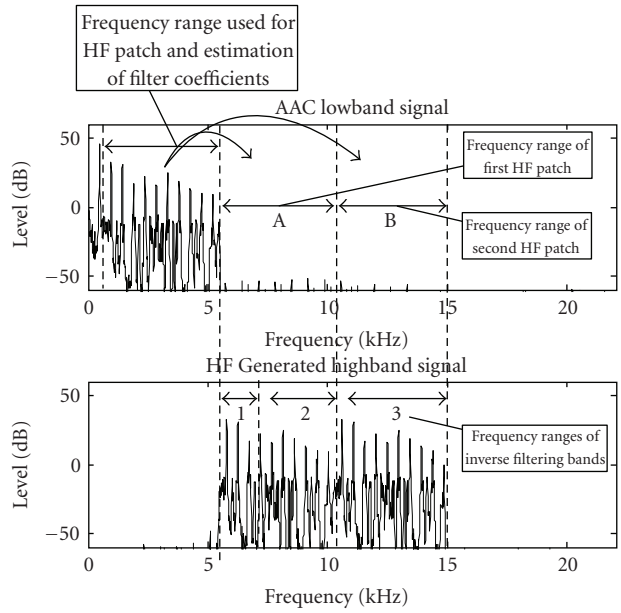

Figure 10: Example of high-frequency generation and inverse filtering. The figure shows the frequency spectra of the low-band signal and the subsequent high-band signals. The signal is an excerpt of a classical music piece coded at 24 kbps mono. The frequency range for SBR, as given in the bit stream for the configuration used, is 5.5 kHz to 15 kHz. This range is covered by two consecutive patches, A and B, where A has a larger frequency range. Finally, three inverse filtering regions are given by the bit stream, where the frequency border of the second and third region coincide with the patch border.

region. Thus, the suitable prediction error filter coefficients are available for all subbands within region A and B. Hence, for all the QMF subbands within the region 1 in Figure 10 an inverse filtering is done within each subband, given the corresponding prediction error filter estimated on the corresponding low-band subband samples and the chirp factor signaled in the bit stream for the specific region.

It should be noted that all the processing done in the HF Generation module is done frame-based on a time segment indicated by the outer borders of the SBR frame.

The generated high-band signals are subsequently fed to the envelope adjusting unit.

Envelope Adjustment. The most important, and also the largest part of the SBR data stream, is the spectrotemporal envelope representation of the high-band. This envelope representation is used to adjust the energy of the generated high-band subband signals. The envelope adjusting unit first performs an energy estimate of the high-band signals. An accurate estimate is possible because of the complex-valued subband signal representation. The resulting energy samples are subsequently averaged within segments according to control signals from the data stream. This averaging produces the estimated envelope samples. Based on the estimated envelope and the envelope representation extracted from the data stream, the energy of the high-band subband samples in the respective segments are adjusted.

As previously outlined sinusoids present in the original high-band signal that have no corresponding sinusoid in the generated high-band are synthesized in the decoder, and random white noise is added to the high-band signal to compensate for diverging tonal-to-noise ratios of the high-band and low-band.

A noise floor level Q is used to derive the level of noise to be added to the recreated high-band signal, it is defined as the energy ratio between the HF-generated (by means of patching in the HF generator) signal energy and the noise signal energy of the final output signal.

Given the calculated gain values, a limiting procedure is applied. This is designed to avoid the need to excessively high-gain values due to large differences in the transposed signal energy and the reference energy given by the original input signal. The limiter is operative to limit high narrow-band gain values while ensuring that the correct wide-band energy is maintained.

QMF Synthesis. The generated high-band signals and the delay-compensated (resulting from the HF generation process) low-band signals are finally supplied to the 64-channel synthesis filter bank, which usually operates at the sampling frequency of the original signal. The synthesis filter bank is just like the analysis filter bank complex-valued, however the imaginary part of the output signal is discarded. Thus, the filter bank generates a real-valued full bandwidth output signal having twice the sampling frequency of the core coder signal.

2.2.4. Other Aspects

Low Power SBR. The SBR tool as outlined in the previous sections is defined in two versions: a High Quality Version and a Low-Power version. The main difference is that the Low-Power version utilizes real-valued QMF filter banks, while the High Quality version utilizes complex-valued filter banks. In order to make the SBR Tool work in the real-valued domain, additional tools are included that strive to minimize the introduction of aliasing in the SBR processing. The main feature is an aliasing detection algorithm that identifies adjacent QMF subbands with strong tonal components in the overlapping range. The detection is done by studying the reflection coefficient of a first-order in-band linear predictor. By observing the signs of the reflection coefficients for adjacent subbands, the subbands prone to introduce aliasing can be identified. For the identified subbands restrictions are put on how much the gain adjustment is allowed to vary between the two subbands.

The following text and figures provide an example of low-power SBR. Envelope adjustment in a real-valued QMF filter bank is displayed in Figure 11.

The upper panel of the Figure 11 illustrates a high-resolution frequency analysis of the input signal superimposed on a stylized visualization of the QMF frequency response. In the middle panel the gain values to be applied

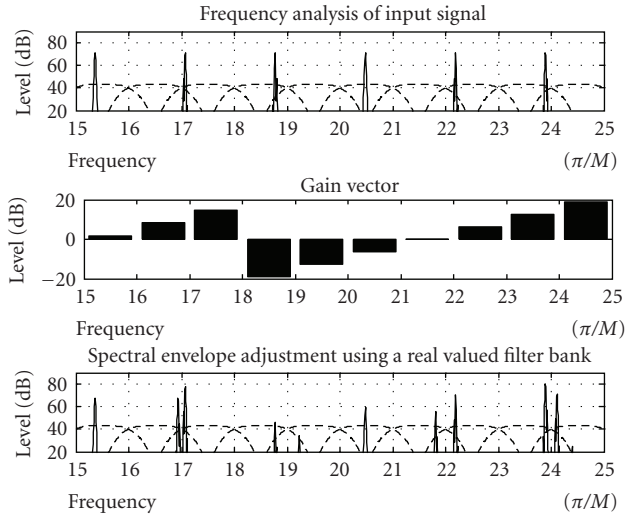

Figure 11: Envelope adjustment in a real-valued QMF filter bank. In the top panel, sinusoids are displayed within QMF subbands. The QMF subband responses are stylistically drawn and at a higher frequency resolution than that of the filter bank, in order to illustrate where within the QMF subband the sinusoids are located. The second panel illustrates the gain vector as calculated by the envelope adjustment module, where the gain values are given for every subband. The third panel illustrates the output after envelope adjustment. Here it is evident that, for example, the sinusoid located between subbands 16 and 17, where also the gain values differ between the subbands, will produce an aliasing component that is not part of the signal in the top panel.

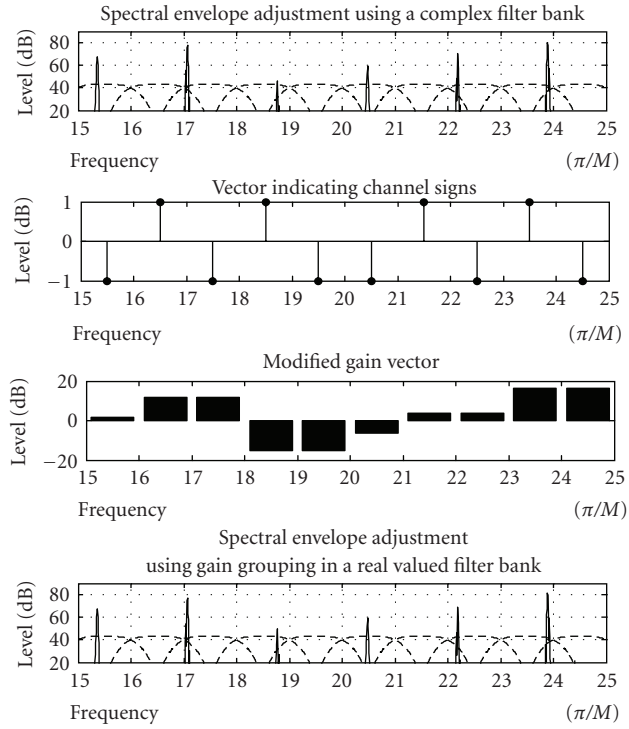

Figure 12: Envelope adjustment in a real-valued QMF filter bank. In the top panel, sinusoids are again displayed within QMF subbands. The second panel illustrates the signs calculated for the different subbands as a function of the reflection coefficients of the subbands. As is clear from the figure, a lower subband with sign 1 adjacent to a higher subband with sign -1 indicates that the two subbands have a shared sinusoid in the overlapping range. In the third panel, the modified gain vector is displayed. Here it is clear that the gain-values for the subbands that share a sinusoid in the overlapping range are identical. The lower panel illustrates the output after envelope adjustment, and as can be seen no aliasing is introduced due to the gain adjustment.

on every subband are displayed. As can be seen these vary from subband to subband. In the bottom panel the high-resolution frequency analysis is again displayed, albeit this time after application of the gain values. As can be observed from the figure, aliasing is introduced.

Figure 12 demonstrates aliasing detection and aliasing reduction. This figure is very similar to Figure 11 except for a new panel with "channel signs." These signs are derived from the reflection coefficients of a first-order predictor, where

$$\text{sign}(x) = \begin{cases} (-1)^k & \text{if } \alpha_1 < 0, \\ (-1)^{k+1} & \text{if } \alpha_1 \geq 0, \end{cases} \tag{4}$$

and where α_1 is given by the prediction error filter

$$A(z) = 1 - \alpha_1 z^{-1} \tag{5}$$

obtained by in-band linear prediction of the subband samples, and k indicates the subband (indexed from zero). Given the definition of the signs and certain relations between the signs of adjacent subbands, the reduction of aliasing can be established by modifying the gain values in the gain vector. For adjacent subbands where the lower subband (in frequency) has a positive sign, and the higher subband (in frequency) has a negative sign, the gain values must be calculated dependently. For all other situations the gain values for the adjacent subbands can be calculated independently. As can be seen from the bottom panel of Figure 12, the use of this algorithm avoids aliasing.

Downsampled SBR. It has been made clear in the previous sections that the combination of AAC and SBR is a dual-rate system. This means that the sampling rate of the output signal from the HE-AAC decoder will always be twice that of the sampling rate of the underlying AAC decoder. Hence, for a normal operation point the AAC will operate at 24 kHz, while the SBR Tool operates at 48 kHz. The dual-rate operation is evident from Figure 13.

For some situations it may be desirable to have an output sampling frequency that is the same as that of the core coder (AAC). One reason is complexity, since for some scenarios, a lower sampling rate output may be desired due to the costs of having D/A converters supporting high sampling rates. This is achieved by operating the SBR Tool in a downsampled mode. When the HE-AAC decoder is operated in the downsampled mode, the synthesis filter bank at the final stage of the SBR decoder is modified. The 64 band QMF synthesis filter bank is replaced by a 32 band QMF synthesis filter bank processing only the lower half of the spectrum of the combined AAC and SBR signal. The result

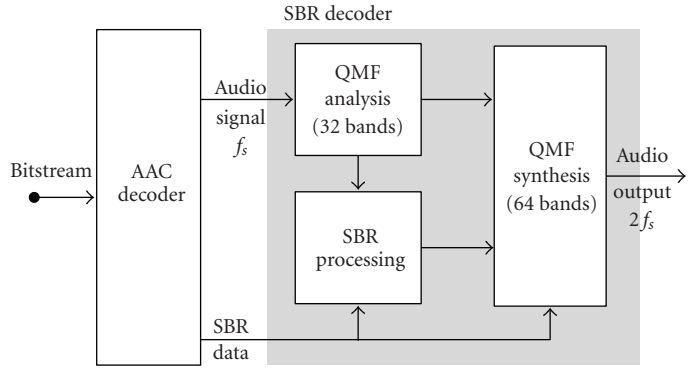

Figure 13: Dual rate structure of the HE-AAC decoder.

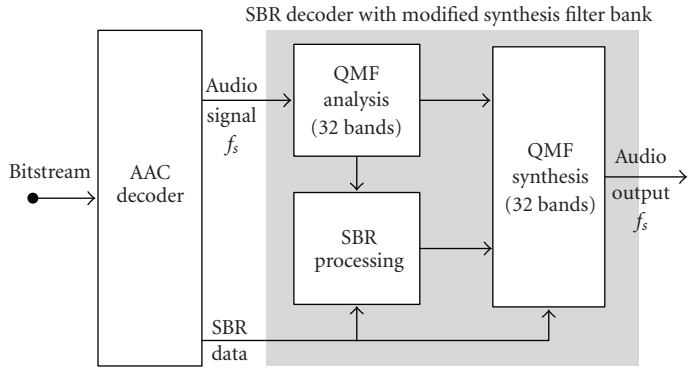

Figure 14: Modified HE-AAC decoder operating in downsampled mode.

is equivalent to operating the decoder in the normal dual-rate decoder, followed by LP-filtering and 1/2 rate downsampling. Apart from the modification of the synthesis filter bank, the remainder of the HE-AAC decoder is left unchanged. This is displayed in Figure 14.

Apart from the application where a low sampling rate output is desired due to complexity constraints, the downsampled SBR mode also serves another purpose. When scaling towards higher bit rates it may be desirable to run the AAC core coder at a higher sampling frequency, for example, 44.1 kHz. Hence, an SBR encoder can operate on a 44.1 kHz input signal, and upsample the signal in the encoder to 88.2 kHz, thus enabling the dual-rate mode. The SBR decoder subsequently operates on the 44.1/88.2 kHz dual-rate signal, but does so in a downsampled mode, ensuring that the output signal has the 44.1 kHz sampling rate equal to that of the original input signal. More information on sampling rate modes in High Efficiency AAC is given in [16].

Scalable Systems. For certain applications scalable systems may be of interest. Scalable in this context refers to a data stream where different information is put in different layers of the stream and, depending on reception conditions, a decoder can choose how many of the layers it decodes. As an example, a base layer or lower layers in the stream may have a higher amount of error protection, while higher layers may not, hence requiring better reception conditions in order to allow decoding. Examples of these kinds of scalable systems using SBR include Digital Radio Mondiale (DRM). The use of SBR as an additional bandwidth extension tool for an underlying core coder lends itself very well to scalable systems. One common way of achieving scalability with waveform codecs is to vary the audio bandwidth depending on the available layers. If only the core layer is available, the output signal has a reduced bandwidth, and when additional layers are available the bandwidth of the output signal is increased. The downside of this approach is that it can be highly annoying to listen to a signal with varying audio bandwidth. Since SBR is a bandwidth extension tool it is the perfect solution for this problem. When SBR is combined with a scalable core codec such as AAC Scalable, the SBR information is put in the core layer. The SBR bit stream comprises data that enables to reconstruct the maximum amount of SBR bandwidth used for any of the layers in the stream. Hence, even if the only the lowest layer is available, the output signal will have full audio bandwidth. If higher layers are available, parts of the SBR frequency range will be replaced by waveform coded segments obtained from decoding the enhancement layer with the underlying core coder. This process is illustrated in Figure 15.

In the top left panel of Figure 15 a spectrum of the two AAC layers (the core layer AAC_0 and the enhancement layer AAC_1) is given. In the top right of the figure, the

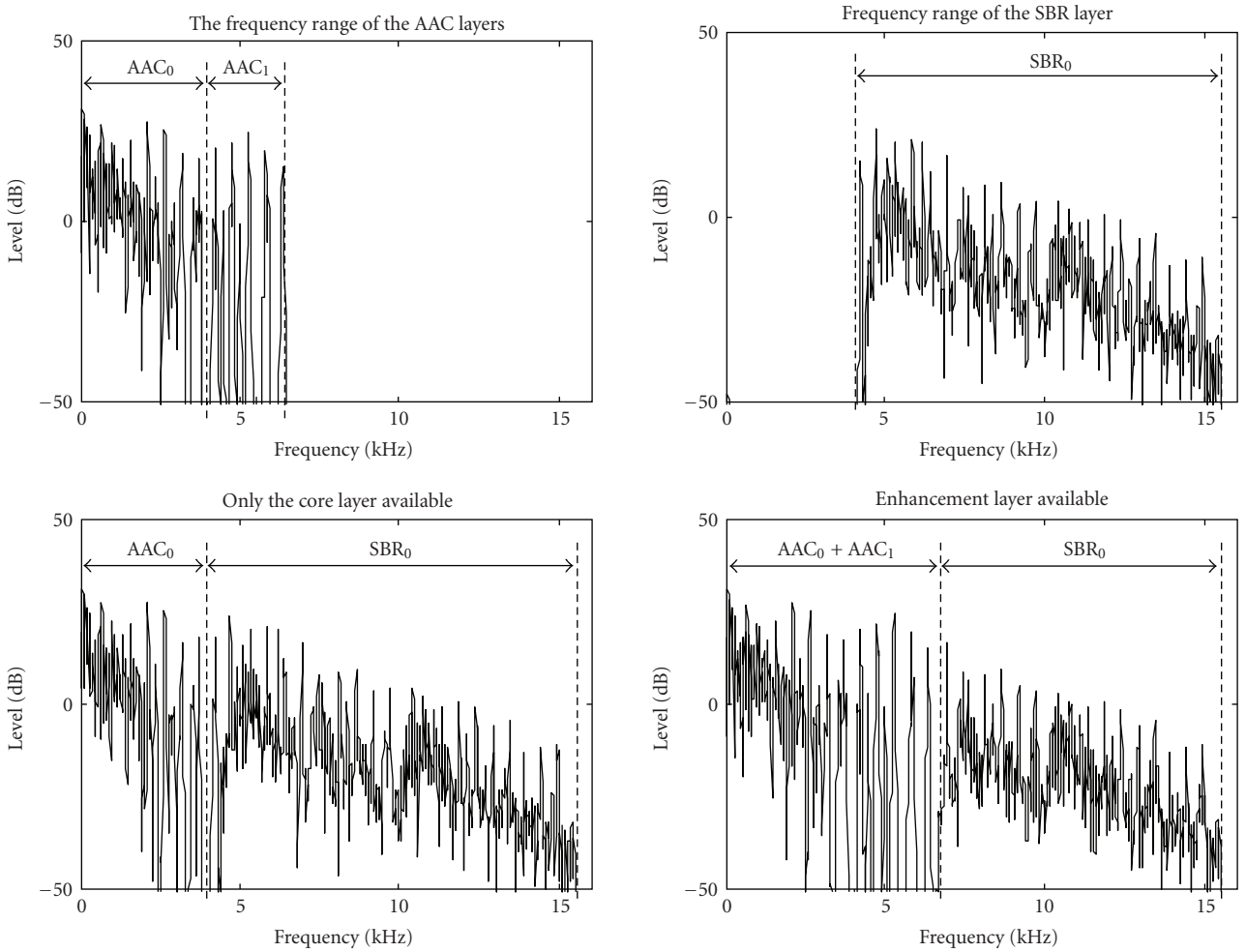

Figure 15: Illustration of scalability. The panels display the frequency ranges and the spectral content of the core layer and the first enhancement layer of a scalable AAC + SBR bit stream. The bit stream contains 3 layers, the first being a 20 kbps monolayer, the second layer adding a 16 kbps enhancement making it in total a 36 kbps mono bit stream. The layers are indicated by the subscript where AAC_0 is the core layer, and AAC_1 is the first enhancement layer. The SBR data is stored in the core layer, and thus labeled SBR_0.

frequency range that can be recreated using the SBR data stored in the core layer is displayed, and a spectrum of the SBR signal available for this range is shown. It is clear that the SBR information covers the widest frequency range required for any combination of layers. In the bottom left figure, the bandwidth relation of the core coder and the SBR tool is illustrated for the scenario where only the core layer is available. In the bottom right figure, the bandwidth relation of the core coder and the SBR tool is illustrated for the scenario where the core layer and the first layer is available. As can be seen from the bottom right picture, the lowest part of the SBR range has been replaced by the core coder.

Apart from supporting bandwidth scalable core coders, the SBR tool can also work in conjunction with mono to stereo scalability. This means that the SBR data can be divided into two groups, one group representing the general SBR data and level information of the one or two channels, and the other group representing the stereo information. If the core coder employs mono/stereo scalability, that is, the base layer contains the mono signal, and the enhancement layer contains the stereo information, the SBR decoder can apply only the monorelevant SBR data to a mono signal and omit the stereo specific parts if only a monocore coder signal is available. If the enhancement layer is decoded, and the core coder outputs a stereo signal, the SBR tool operates on the stereo signal as normal using the complete SBR data in the stream.

MPEG-2 Bit Streams. Although the focus of the present paper is on the MPEG-4 version of SBR, it should be noted that the exact same tool is standardized in MPEG-2 as well. Hence, the MPEG-2 AAC and SBR combination is also defined. This is important for certain applications relying on MPEG-2 technology while still wanting to achieve state-of-the-art compression by using SBR in combination with AAC.

2.3. Listening Tests. At the end of the two-year standardization process a rigorous verification test was performed. Two types of tests were done, a (MUlti Stimulus test with

Table 1: Codecs under test.

Coding scheme	Label	Bit rate (mono/stereo)	Sampling rate (kHz)	Typical audio bandwidth (kHz)
MPEG-4 AAC profile	AAC 48/60 kbps	48/60 kbps	32	10/13.5
MPEG-4 HE-AAC	HE-AAC 32/48 kbps	32/48 kbps	24/48	15.5
Anchors and reference	Hidden reference	16-bit PCM stereo	48	24
Anchors and reference	Anchor 3.5 kHz	16-bit PCM stereo	48	3.5
Anchors and reference	Anchor 7 kHz	16-bit PCM stereo	48	7.0

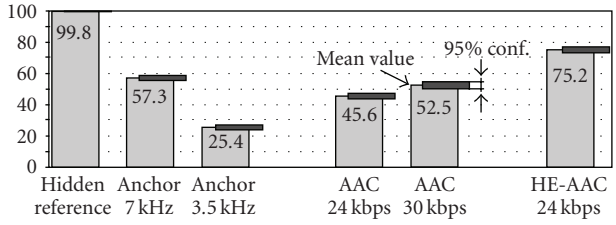

Figure 16: Listening test results for mono MUSHRA tests. The scores are the average scores over all items and test-sites (adapted from [19]).

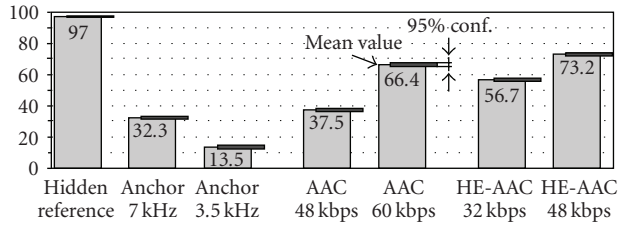

Figure 17: Listening test results for stereo MUSHRA tests The scores are the average scores over all items and test-sites (adapted from [19]).

Hidden Reference and Anchor) MUSHRA test [17] and a (Comparative Mean Opinion Score) CMOS test [18]. The MUSHRA test compared the performance of MPEG-4 HE-AAC with that of MPEG-4 AAC when coding mono and stereo signals at bit rates in the range 24 kbps per channel, while the CMOS test was used to show the difference between High Quality SBR and Low Power SBR. Two test sets were selected, one for mono testing, and one for stereo testing. The items were selected from 50 potential candidates by a selection panel identifying ten items considered critical for all of the systems under test.

The codecs under test for the verification tests are outlined in Table 1. The listening tests were performed at France Télécom, T-Systems Nova, Panasonic, NEC, and Coding Technologies.

The listening test results are presented in Figures 16 and 17. From the listening tests it is clear that the SBR enhanced AAC technology (High Efficiency AAC Profile) performs better than the MPEG-4 AAC Profile when the latter is operating at a 25% higher bit rate (i.e., 30 versus 24 kbps for mono, and 60 versus 48 kbps for stereo).

The SBR technology in combination with AAC as standardized in MPEG under the name High Efficiency AAC (also known as aacPlus) offers a substantial improvement in compression efficiency compared to previous state-of-the-art codecs. It is the first audio codec to offer full bandwidth audio at good quality at low bit-rate. This makes it the ideal codec (and enabler) for low bit-rate applications such as Digital Radio Mondiale and streaming to mobile phones.

3. MPEG-4 SSC

3.1. Parametric Mono Coding. Current standardized and proprietary coding schemes are primarily build based on waveform coding techniques. These coding algorithms translate the incoming signal to the frequency domain by use of a subband or transform technique. Furthermore, a psychoacoustic model analyzes the incoming signal as well and determines the number of bits for quantization of each of the subband or transform signals. For an overview, see [20].

The subband or transform audio coding schemes primarily exploit the destination (human ear) model; the psychoacoustic model tells us where signal distortions (quantization) are allowed such that these are inaudible or least annoying. In speech coding, on the other hand, source models are primarily used. The incoming signal is matched to the characteristics of a source model (the vocal tract model), and the parameters of this source model are transmitted. In the decoder, the source model and its parameters are used to reconstruct the signal. For an overview on speech coding, please refer to [21].

The speech coding approach guarantees that the reproduced signal is in accordance with the model. This implies that if the model is an accurate description, the generated signal will sound like stemming from a vocal tract and will therefore sound natural though not necessarily identical to the incoming signal.

For audio, it is not possible to directly follow an approach like in speech coding. There are many sources in audio and these have quite different characteristics. The consequences of using a too restrictive source models can be devastating to the sound quality. This is already demonstrated by speech coders operating at low bit-rates; input signals other than speech typically result in a poor quality of the decoded output signals.

Nevertheless, a model is used in parametric coding. This is called a signal model to distinguish it from source models as are used in speech coding. The origin of the signal model is more based on destination properties

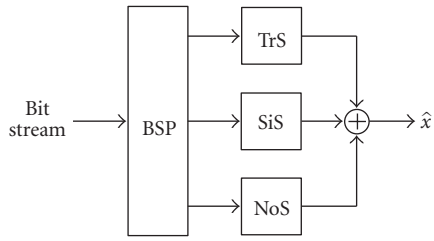

Figure 18: Decoder scheme producing the decoded signal $\hat{x}$ from the bit stream. The decoding consists of a bit stream parser (BSP), a transient synthesizer (Trs), a sinusoidal synthesizer (SiS), and a noise synthesizer (NoS).

(i.e., the human hearing system) in the sense that it tries to describe perceptually-relevant acoustic events. Consequently, parametric coding is also related to musical synthesis. However, the distinction between source and destination models is arguable; for example, many musical instruments create tonal components and biological evolution presumably leads to a tight connection between destination and source characteristics.

The promises that the parametric approach holds are therefore as follows. First of all, the signal model should always lead to an impression of an agreeable sound even at low bit rates. Thus a graceful degradation of sound quality with bit rate should be feasible. This is a property which is difficult to attain in conventional audio coding techniques. Secondly, since the idea is to model acoustic events, we may be able to manipulate these events (like in musical synthesis), a feature clearly not feasible in conventional audio coding.

At various universities, prototype parametric audio coders have been developed [22–29]. Prior to the parametric coder described in this paper, there was only one standardized parametric audio coder: HILN [30] in MPEG-4 Audio Version 2 [31].

In the Sinusoidal Coder (SSC) that is described here and which is standardized in MPEG-4, three objects can be discerned. The first one comprises tonal components. These are modeled by sinusoids. This idea seems to be originated from speech coding [32–34]. The second one is a noise object. Also this object is present in speech coders, only there segments are typically denoted as either voiced or unvoiced, corresponding to noise and periodic excitations. In audio, an early reference to simultaneous use of sinusoidal and noise coding is [35].

Both sinusoidal modeling and noise modeling assume that the signal segment being modeled is stationary. In view of bit rate and frequency resolution, these segments may not be too short. Consequently, one can find audio segments that, given the analysis segment length, contain clearly instationary events. A famous example forms the castanets excerpt, which is therefore a critical item for almost any coder. In view of this, it was decided to introduce a third object which is the transients. The coder not only uses a separate transient object but also adapts the windowing for the sinusoidal and noise analysis and synthesis on basis of detected transients.

3.1.1. SSC Decoder

Overview. The SSC decoder is depicted in Figure 18. As described in the previous section, the idea is that a mono audio signal can be described by three basic signal components: transients, sinusoids, and noise. The information on these components is contained in the bit stream and the decoder uses a parser Bit Stream Parser (BSP) to split this stream. The three basic signal components are decoded using a transient, sinusoidal, and noise synthesizer (TrS, SiS, and NoS, resp.). Adding these signals gives a decoded mono audio signal ($\hat{x}$).

A detailed description of the decoder can be found in the MPEG-4 document [36]. In the following, we merely outline the operations of the different modules.

Transient Synthesis. The bit stream contains transient information. First of all, transient positions are transmitted together with a type parameter. There are two types: a step-like transient and a Meixner transient. In both cases, the transient position is used to generate adapted overlap-add windows for the sinusoidal and noise synthesis. Thus this information is shared by the three synthesizers TrS, SiS, and NoS.

In the case of a Meixner window, a Meixner envelope is created and multiplied by a number of sinusoids thus defining a transient phenomenon [37–39]. The discrete-time Meixner envelope is given by

$$g(n) = \left(1 - \xi^2\right)^{b/2} \sqrt{\frac{(b)_n}{n!}}\, \xi_n, \tag{6}$$

with $b > 0$, $0 < \xi < 1$ and $n = 0, 1, \ldots$. The parameters b and ξ define the rise and decay time of the transient envelope. In case of a step-like transient, no signal is created by the transient generator. However, due to the use of the adapted overlap-add windows in the sinusoidal and noise synthesizers, a transient phenomenon is created in the mono signal $\hat{x}$ for the step transient as well.

Sinusoidal Synthesis. The sinusoidal data is contained in so-called sinusoidal tracks. From these tracks, information on the number of sinusoids, their frequencies, amplitudes, and phases is available for each frame. These signals are generated to produce a waveform per frame. Typically, the frames are overlap-added using an amplitude-complementary Hanning window with 50% overlap. In case of a transient, fade-in or fade-out of these overlap-add windows are shortened and positioned around the pertinent transient position.

Noise Synthesis. The noise synthesizer consists of a white noise generator with unit variance. The bit streams contains data concerning the temporal envelope per 4 frames. The envelope is generated [39] and applied to the noise. Next, this temporally shaped noise is an input to a linear prediction synthesis filter based on Laguerre filter [40]. The data on the filter coefficients are contained in the bit stream per frame. The generated noise is overlap added using power complimentary windows.

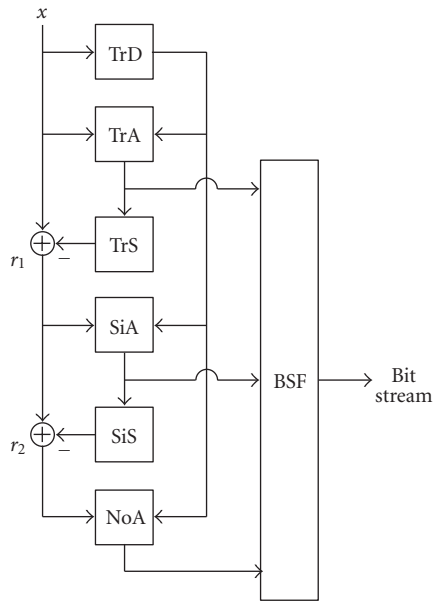

Figure 19: Encoder scheme producing a bit stream from an audio signal. It consists of a transient detector (TrD), a transient analyzer (TrA), a sinusoidal analyzer (SiA), a noise analyzer (NoA), and a bit stream formatter (BSF).

3.1.2. SSC Encoder

Overview. The SSC encoder is not standardized by MPEG and as such several designs are possible. We will discuss the structure of the encoder we developed and different possible mechanisms within this structure.

The mono encoding scheme (Figure 19) implements the opposite process to the decoder in a cascaded manner. The coder analyzes the input signal x and describes it as a sum of three basic components. To this end, it uses a transient detector (TrD) which detects transients and estimates their starting position. This information is fed to the transient analysis (TrA) which estimates the transient component parameters and feeds these to a transient synthesizer. In the transient synthesizer (TrS), the estimated waveform captured in the transient parameters is generated and subtracted from the input signal, thus making a first residual r_1.

The first residual is an input to a sinusoidal analyzer (SiA) which also uses the estimated transient positions. This information is exploited in order to prevent measuring over nonstationary data which is done by adaptation of the analysis windows. The sinusoidal parameters are fed to a sinusoidal synthesizer (SiS) which generates a waveform. This waveform is subtracted from the first residual signal thus generating a second residual signal r_2.

The signal r_2 is fed to a noise analyzer (NoA). This analyzer tries to capture the spectral and temporal envelopes of the remaining signal ignoring its specific waveform. Also in this analysis module, the transient position estimates are used for window adaptation.

The parameter streams generated by the transient detector and the various analysis stages are fed to a bit stream formatter (BSF). At this stage, irrelevancy and redundancy of the parameter streams are exploited and the data is quantized. The quantized data is stored in a bit stream.

Though the concept of separation in these three different objects is similar to the work presented in [41], there are large differences between the approaches. This holds for the different models which are used for the noise and transient components, but also in the sense that [41] subdivides the input signal in time-frequency tiles where each tile is exclusively modeled by one of the three components.

Transient Analysis. The transient analysis is only performed when the transient detector signals the occurrence of a sudden change in the input signal. The detector can be build on basis of detection of changes of energy [42] where these changes are defined over the entire frequency range or over different frequency bands. Next to detection of a transient, the detector estimates the start position of the transient.

When the transient detector signals the occurrence of a transient in a frame, the transient analysis module becomes active. On basis of the input signal and the received transient start position, it first determines the character of the transient. If the transient phenomenon is shorter than the analysis frame lengths used in the sinusoidal and noise analysis (typically in the order of tens of milliseconds), a Meixner modeling stage becomes active. Otherwise, the transient is designated as a step transient and no separate modeling is applied. Instead, the transient position information is used in the sinusoidal and noise analysis for window adaptation.

For a short transient phenomenon, the Meixner modeling stage is employed. It determines a time-domain envelope and a number of sinusoids underneath the envelope. For a detailed description of the time-domain envelope modeling process, we refer to [37–39]. This transient is subtracted from the input signal in order to ensure that this intra-frame transient is removed as much as possible before entering the sinusoidal and noise analysis, since these stages operate under the assumption that the input signal is quasistationary.

Sinusoidal Analysis. Sinusoidal analysis is a well-known technique for which many algorithms exist. Of these we mention peak-picking, matching pursuit, and psychoacoustic weighted matching pursuit. Whatever method is used, a set of frequencies, amplitudes, and phases evolves as outcome. Extended models including amplitude and frequency variations [43, 44] for more accurate signal modeling have been proposed as well but are not used in the SSC coder.

In contrast to the HILN coder, SSC does not use harmonic complexes as an object. Though a harmonic object can act as a compaction of the sinusoidal data, it was decided not to use for several reasons. Firstly, harmonic complexes need to be detected which may involve wrong detection decisions. Secondly, the linking process becomes more complicated because linking has to be established not only between sinusoids and harmonic complexes separately but also in between these two. Lastly, the signaling of links between harmonic and individual sinusoids would lead to a much more complex structure of the bit stream.

The sinusoids from subsequent frames are linked in order to obtain sinusoidal tracks. Transmission of track data is relatively efficient since the characteristic property of a track is the slow evolution of the sinusoidal amplitude and frequency. Only the phase has a more complicated character. In principle, the phase can be constructed from the frequency since these are related by an integral relation. Thus in order to arrive at low bit rate sinusoidal coders, the phase is typically not transmitted. However, phase is an important property: phase relations between different tracks are relevant for the perception and can be severely distorted when not transmitting the phase. Therefore, a new phase transmission mechanism was conceived which transmits the unwrapped phase and thus implicitly the frequency parameter as well [45]. This is slightly more expensive in terms of bits than discarding the phase but improves the perceived quality and is much more efficient than separate frequency and phase transmission.

In order to remain within a predefined bit budget, the estimated sinusoids are typically ordered in importance and the number of transmitted sinusoids is reduced when necessary. An overview of methods for doing so can be found in [46].

Noise Analysis. The noise analysis characterizes the incoming signal by two properties only: its spectral shape (spectral envelope) and its temporal envelope (power over time). As such, the analysis consists of two distinct stages. First, the spectral envelope is extracted. The spectral envelope is obtained by using linear prediction based on the Laguerre systems [40]. The use of these filters is motivated by the fact that it allows modeling of spectral details in accordance with their relevance on a Bark frequency scale [47].

The resulting spectrally flattened signal is analyzed for its temporal structure. This structure is analyzed over several frames simultaneously in order to obtain a good balance between required bit rate and modeling capability. The envelope modeling is done by linear prediction in the frequency domain [48, 49].

Since both linear prediction stages yield normalized envelopes, a separate gain parameter is determined and fed to the BSF as well.

3.1.3. SSC Bit Stream

Overview. The bit stream formatter receives the data from the analyzers and puts them with headers into a bit stream. Details of the bit stream defined are described in [36]. We will consider the main data only.

The transient data comprises the transient position, transient type, envelope data, and sinusoids. The transient position and type are directly encoded. The envelopes are restricted to a small dictionary. The sinusoids underneath the envelope are characterized by their amplitude, frequency, and phase. Amplitude and frequency quantization can be done with different levels of accuracy. The amplitudes are uniformly quantized on a dB scale with at least 1.5 dB accuracy. The frequencies are uniformly quantized on an ERB scale [50]. For a 1 kHz frequency the accuracy is at least 0.75%. Both amplitude and frequency are Huffman encoded. The phases are encoded using 5 bit uniform quantization.

The sinusoidal data comprises sinusoidal tracks. This can be divided in start data and track data, that is, everything after the start of a sinusoid until and including its death. The start data are sorted according to ascending frequency, quantized uniformly on an ERB scale and differentially encoded. The amplitude data is sorted in correspondence with the frequencies, uniformly quantized on a dB scale and differentially encoded using Huffman tables. The accuracy of both the amplitudes and frequency quantization can be set to different levels. The start phases are encoded using 5 bits.

The sinusoidal track data consists of unwrapped phases and amplitudes. The unwrapped phase data along a track is a combination of the originally estimated frequency and phase per frame and those from the previous frame (as established by the linking). This unwrapped phase data is input to a 2-bit ADPCM mechanism [45]. The amplitudes are quantized on a dB scale and differentially encoded along a track using Huffman coding.

The noise data consists of three parts: a gain, a spectral, and a temporal envelope. The gain is quantized uniformly on a dB scale and Huffman encoded. The prediction coefficients describing the spectral envelope are mapped onto Log Area Ratios (LARs) and quantized with an accuracy according to index number. The prediction coefficients describing the temporal envelope are mapped to Line Spectral Frequencies (LSFs) and quantized.

Most of the data is updated every 384 samples for 44.1 kHz input signal, other data has an update being a multiple of this. The update of 384 samples corresponds to a subframe. Eight consecutive subframes are stored into one frame of the bit stream.

3.2. Parametric Stereo Coding. Since most audio material is produced in stereo, an efficient coding tool should also exploit the redundancies and irrelevancies of both channels simultaneously. Since it is not straightforward to use standard stereo coding tools like mid/side stereo [51] and intensity stereo [52] in conjunction with parametric coding, and since the aim also was to develop a general stereo coding tool for low bit rates, the novel Parametric Stereo (PS) tool was developed where the stereo image is coded on the basis of spatial cues. The PS tool as standardized in MPEG was developed in 2003 and primarily aimed to enhance the performance of SSC and HE-AAC at low bit rates.

In the context of SSC, the spatial cues can be considered to form the fourth object. Here, we treat this issue separately since, basically, this coding tool can be used in conjunction with any mono coder. Depending on the mono coder, it may be worthwhile to integrate the PS tool with parts of the mono coder. This has been done with HE-AAC; by sharing infrastructural parts like the time/frequency transform, a lean implementation is enabled with great savings in complexity. Details on the combination of HE-AAC an PS can be found in Section 4, and more theoretical background on the PS tool is available in [53, 54].

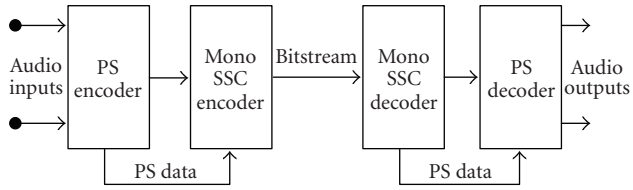

Figure 20: Structure of the SSC encoder (left) and decoder (right) extended with PS. The PS encoder generates a down-mix and PS parameters. The resulting down-mix is subsequently encoded using a mono SSC encoder. The resulting mono bit stream and the PS parameters are combined into a single output bit stream. At the decoder side, the mono SSC decoder generates a time-domain down-mix signal, which is converted to stereo by a PS decoder based on the transmitted PS data.

3.2.1. Stereo Analysis

Overview. The PS encoder proceeds the SSC encoder (see Figure 20). The PS encoder compares the two input signals (left and right) for corresponding time/frequency tiles. The frequency bands are designed to approximate the psychoacoustically motivated ERB scale, while the length of the segments is closely matched to known limitations of the binaural hearing system (see [53, 54]). Essentially, three parameters are extracted per time/frequency tile, representing the perceptually most important spatial properties.

(i) Interchannel Level Difference (ILD), representing the level difference between the channels similarly to the "pan pot" on a mixing console.

(ii) Interchannel Phase Difference (IPD), representing the phase difference between the channels. In the frequency domain this feature is mostly interchangeable with an Interchannel Time Difference (ITD). The IPD is augmented by an additional Overall Phase Difference (OPD), describing the distribution of the left and right phase adjustment.

(iii) Interchannel Coherence (ICC), representing the coherence or cross-correlation between the channels.

While the first two parameters are coupled to the direction of sound sources, the third parameter is more associated with a spatial diffuseness (or width) of the source.

Subsequent to parameter extraction, the input signals are down-mixed to form a mono signal. The down-mix can be made by trivial means of a summing process, but preferably more advanced methods incorporating time alignment and energy preservation techniques are incorporated to avoid potential phase cancellation (and hence resulting timbre changes) in the down-mix. The down-mix is subsequently encoded using a mono SSC encoder resulting in a mono bit stream. The PS data are properly quantized according to perceptual criteria [54], while redundancy is removed by means of Huffman coding. Finally, the mono SSC bit stream is combined with the PS data into a joint output bit stream.

3.2.2. Stereo Synthesis

Overview. The SSC decoder extended with a PS decoder is also outlined in Figure 20 and basically comprises the reverse process of the corresponding encoder. The SSC decoder generates a mono down-mix. Subsequently, the PS decoder reconstructs stereo output signals based on the PS parameters.

The PS decoder is outlined in more detail in Figure 21. The input signal (a mono decoded signal resulting from the SSC decoder) is processed by a hybrid analysis QMF bank. The hybrid QMF analysis bank is the same as used in HE-AAC (in SBR), extended with a second filter step to increase the spectral resolution for low frequencies according to psychoacoustical requirements (cf. [55]). The resulting subband signals are subsequently processed by a decorrelation filter and a mixing stage. The decorrelation filter generates a artificial side signal based on the mono down-mix. The design of the decorrelation process is technically related to artificial reverberators but also includes many PS integration aspects due to, for example, the dynamics of the control parameters. This is thoroughly discussed in [56]. The hybrid QMF-domain output signals are obtained as a certain linear combination of the mono and side signal. This linear combination, referred to as mixing or rotation, is controlled by the PS parameters (ILDs, IPD/OPDs, ICCs). This process of up-mixing the mono signal, $M_{k,i}$ with aid from the decorrelated mono signal, $D_{k,i}$, into the final estimate of left and right signal $(\hat{L}_{k,i}, \hat{R}_{k,i})$ is expressed by

$$\begin{bmatrix} \hat{L}_{k,i} \\ \hat{R}_{k,i} \end{bmatrix} = \mathbf{H}_{k,i} \begin{bmatrix} M_{k,i} \\ D_{k,i} \end{bmatrix} \tag{7}$$

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

$$\mathbf{H}_{k,i} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix}, \tag{8}$$

where k and i denote the frequency subband and the QMF time slot, respectively. The elements in the up-mix matrix, $\mathbf{H}_{k,i}$ are the only up-mix variables actually derived from the stereo parameters. Details about the calculation of these matrix elements can be found in [36, 54, 57]. Finally, two hybrid QMF banks are used to generate the two output signals.

3.3. SSC Performance. In order to be included in the MPEG-4 standard, the developed high-quality parametric coder needed to pass the requirements that were set out at the start of the standardization process. These requirements were twofold.

(i) The coder should provide the same quality in the mean at a 25% less bit rate compared to the existing MPEG-4 state-of-the-art technology.

(ii) The coder should not provide less quality for any item when operating at the same bit rate as existing MPEG-4 state-of-the-art technology.

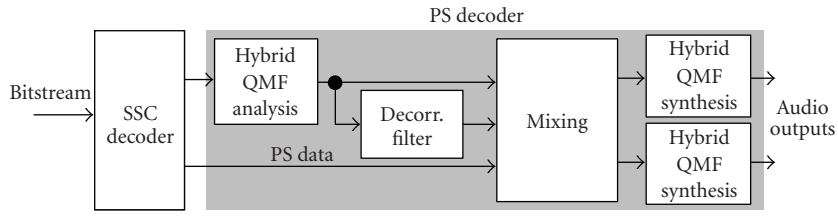

Figure 21: Structure of the QMF-based PS decoder. The signal is first fed through a hybrid QMF analysis filter bank. The filter-bank output and a decorrelated version of each filter-bank signal is subsequently fed into the mixing and phase-adjustment stage. Finally, two hybrid QMF banks generate the two output signals.

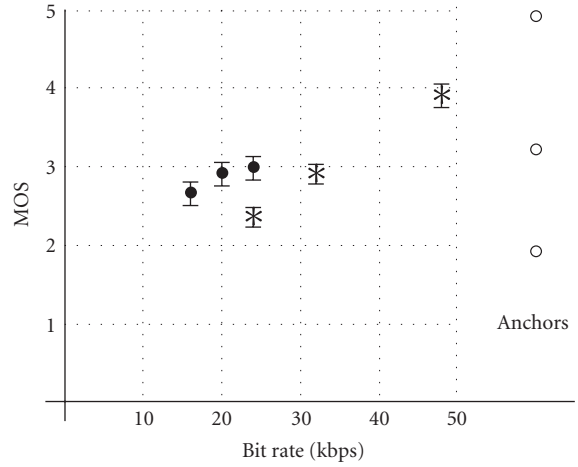

Figure 22: Average scores (MOS) versus bit rate for coded stereo signals. The filled circles indicate the SSC coder (at 16, 20, and 24 kbps), the stars the AAC coder (at 24, 32, and 48 kbps). The open circles on the right-hand side are the anchors (hidden reference, 7 kHz and 3.5 kHz low-pass filtered versions from top to bottom, resp.). The 95% confidence intervals have as typical range 0.3 MOS units. Adapted from [58].

The existing MPEG-4 state-of-the-art technology at that moment in time was AAC.

These requirements have been assessed in a subjective verification test conducted by the Institüt für Rundfunktechnik (IRT). In this section we present some of the results that were reported at that point in time. The data that are discussed are taken from [58].

The listening tests performed for the MPEG-4 standardization were done in two stages. A set of 53 critical items were encoded using the SSC (from Philips) and the AAC encoder (Fraunhofer Gesellschaft, FhG), respectively. The encoding was done at different bit rates and for mono as well as stereo material. The Institüt für Rundfunktechnik (IRT) did a prescreening test to generate a set of 9 critical items to be included in the final listening test which was also executed at IRT. Here, we discuss only the results of this final listening test obtained with the stereo input material, since this is the more relevant data from an application point-of-view. Furthermore, the results from an additional listening test performed at Philips concerning all 53 tests items are presented.

Next to the 9 encoded items, the IRT test included the hidden reference and two anchors, being the original material band-limited at 3.5 and 7 kHz. The test was performed with headphones using 26 listeners. The MUSHRA tool was used and the listeners were instructed to give a Mean Opinion Score (MOS).

The test was supposed to be a blind test. However, since two completely different coding strategies were used, the test was effectively far from blind. The AAC coder reduces its bandwidth when operating at low bit rates and this is always immediately recognized unless there is band-limited material. The SSC encoder never uses a band limitation, this being an ingredient for reaching a high quality encoding. The completely different artifacts introduced by both coding schemes effectively not only prohibit a blind comparison, but also made the ranking of the different coders a complicated task. Also, the results tend to be subject-dependent. For example, in older experiments involving SSC and AAC performed at Philips we found that listeners that are well-acquainted with band-limitations (speech-coding experts) tended to perceive the AAC band limitation as less annoying than most other listeners.

The SSC coder was operated at 16, 20, and 24 kbps, the AAC coder at 24, 32, and 48 kbps. In Figure 22 the results can be found. The crosses indicate the mean score of the AAC coder, the filled circles those of the SSC coder and the open circles on the right reflect the mean of the anchors. The 95% confidence intervals are indicated as well for the AAC and SSC means. Interestingly, the AAC score is dropping rapidly as a function of bit rate, whereas the quality of the SSC coder is not. Comparing AAC at 24 kbps with SSC at the same bit rate, it is clear that the first is statistically significant better. The MOS score for SSC is on the border of a fair and good qualification. According to this test, the score of SSC at 24 kbps is roughly equivalent to that of AAC at 32 kbps. However, in view of the fact that some of the 9 excerpts were rather band-limited [58], we decided to look in somewhat more detail to this comparison.

At Philips, a listening test was performed involving all 53 items for the AAC material at 32 kbps and the SSC material at 24 kbps. The listeners were asked to give a preference and a rating for these data using headphones and the MUSHRA tool. The rating is according to an ITU-R 7-point scale. The results of this test are given in Figure 23. The positive scores 1, 2, and 3 indicate that SSC is slightly better, better and

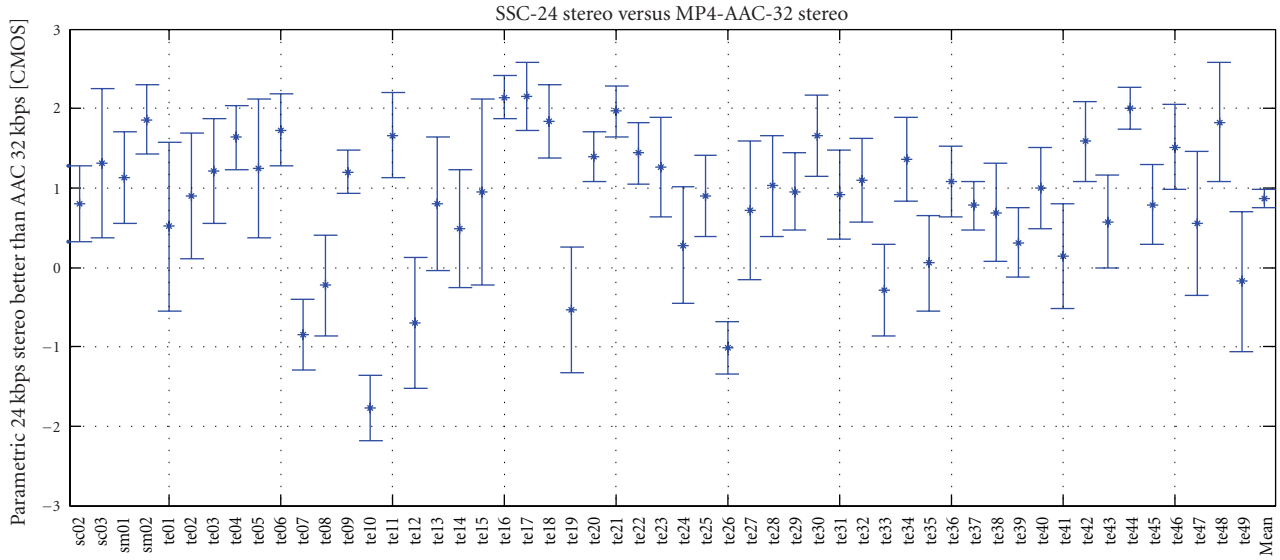

Figure 23: Comparative MOS score for 53 items using SSC at 24 kbps and AAC at 32 kbps for stereo signals.

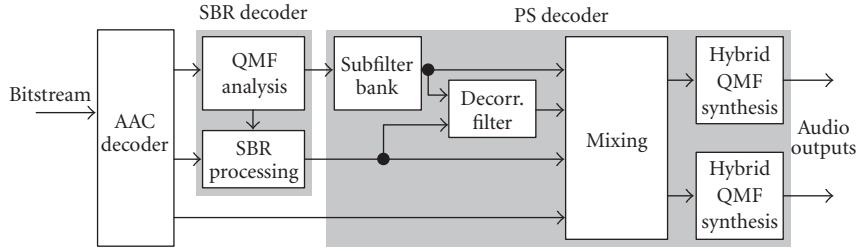

Figure 24: Decoder structure of HE-AAC v2.

much better than AAC, respectively. A negative score has the opposite meaning.

From Figure 23 we see that SSC operating at 25% lower bit rate than AAC (32 kbps) yields a better score for 35 items, a (statistically) equal score for 16 items, and a lower score for only 3 items. We note that exactly these latter three items have been included in the selected items for the test presented earlier and, in that sense, the test presented earlier is rather critical for the SSC coder. If we look at the mean preference over all items, SSC at 24 kbps is rated as slightly better than AAC at 32 kbps, this difference being statistically significant.

4. MPEG-4 HE-AAC v2

4.1. Introduction. The combination of MPEG-2/4 AAC with the SBR bandwidth extension tool, as presented in the Section 2 is also known as *aacPlus* and was standardized in MPEG-4 as HE-AAC [15].

Since the bandwidth extension enabled by SBR is in principle completely orthogonal to the channel extension provided by the Parametric Stereo (PS) tool introduced in Section 3.2, it is of interest to combine both tools in order to utilize the coding gain of both tools simultaneously.

4.2. Combining HE-AAC with PS. When the PS tool presented in this paper is combined with HE-AAC, this results in a coder that achieves a significantly increased coding efficiency for stereo signals at very low bit rates when compared to HE-AAC operating in normal stereo mode. Figure 24 shows a simplified block diagram of the resulting decoder, which is referred to as HE-AAC v2 (or 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 efficient manner directly prior to the final QMF synthesis filter bank. Comparing Figures 13 and 24, it is evident that only the parametric stereo decoding and synthesis, including its hybrid filter bank (here denoted "Subfilter"), have to be added to a mono HE-AAC decoder, plus of course a second QMF synthesis bank. The computational complexity of such a decoder is approximately the same as that of an HE-AAC 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 [55, 57].

The PS tool 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

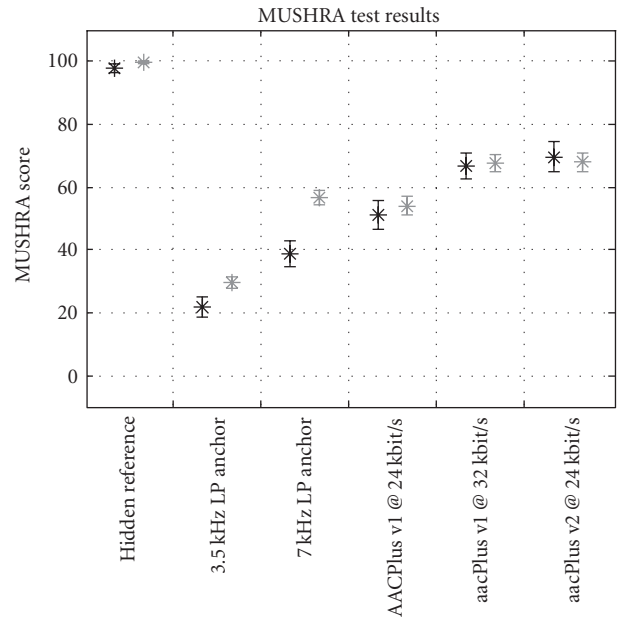

Figure 25: MUSHRA listening test results for two sites (black and gray) showing mean grading and 95% confidence interval for HE-AAC at 24 kbps, 32 kbps, and HE-AAC v2 at 24 kbps (from [59]).

parameter coding and Huffman codebooks, makes it possible to operate this PS system over a large range of bit rates.

When an HE-AAC v2 coder is operated at target bit rate of 24 kbps, the PS parameters require an average side information bit rate of 2 to 2.5 kbps, assuming 20 stereo bands for ILD 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 kbps or more PS side information.

Based on the already existing HE-AAC profile, an HE-AAC v2 profile was defined that, in addition to AAC and SBR, includes the PS tool [60]. In level 2 of the HE-AAC v2 profile, only a "baseline" version of the PS tool is implemented in order to limit the computational complexity. This baseline version includes a simpler version of the hybrid filter bank and does not implement IPD/OPD synthesis, but is still capable of decoding all possible PS bit stream configurations.The HE-AAC v2 decoder implementing level 2 of the HE-AAC v2 profile was also standardized in 3GPP as part of Release 6 [61], where it is referred to as "Enhanced aacPlus."

4.3. Listening Tests. Figure 25 shows subjective results from a listening test comparing HE-AAC using normal stereo coding at 24 and 32 kbps with HE-AAC v2 utilizing the PS tool at 24 kbps [59]. Two sites (indicated in black and gray) participated in this test, with 8 and 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 [17] methodology and included a hidden reference and low-pass filtered anchors with 3.5 and 7 kHz bandwidth.

At both test sites, it was found that HE-AAC v2 at 24 kbps achieves an average subjective quality that is equal to HE-AAC v1 stereo at 32 kbps and that is significantly better than HE-AAC v1 stereo at 24 kbps. It is of interest to relate these results to the MPEG-4 verification test [19]. There, it was found that HE-AAC v1 stereo at 32 kbps achieved a subjective quality that was significantly better than AAC stereo at 48 kbps and was similar to or slightly worse than AAC stereo at 64 kbps. This shows that HE-AAC v2 achieves approximately three times the coding efficiency of AAC for stereo signals. Further MUSHRA tests have shown that HE-AAC v2 achieves a significantly better subjective quality than HE-AAC v1 stereo also for 18 and 32 kbps.

5. Conclusions

An overview of the technology defined in the Amendments 1 and 2 to the 2001 edition of the MPEG-4 Audio standard has been given. The performance of these techniques is discussed on the basis of the delivered audio quality as indicated by listening tests. These show that the SBR, SSC, and PS technologies add so far unreached points in the quality/bit-rate plane. In particular for low bit rate applications the parametric coding techniques constitute valuable tools. This was essentially the basis for the acceptance by MPEG-4.

Since the finalization of the standard, the HE-AAC v2 codec has gained a wide market acceptance and is currently used in several mobile music download services, digital radio broadcasting systems, and Internet streaming applications.

References

[1] Audio Subgroup, "Call for proposals for new tools for audio coding," ISO/IEC JTC1/SC29/WG11 N3794, 2001.

[2] J. Makhoul and M. Berouti, "Predictive and residual coding of speech," *The Journal of the Acoustical Society of America*, vol. 66, no. 6, pp. 1633–1640, 1979.

[3] J. Makhoul and M. Berouti, "High frequency regeneration in speech coding systems," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '79)*, vol. 4, pp. 428–431, Washington, DC, USA, April 1979.

[4] J. Epps and W. Holmes, "A new technique for wideband enhancement of coded narrowband speech," in *Proceedings of IEEE Workshop on Speech Coding*, pp. 174–176, Porvoo, Finland, June 1999.

[5] L. Liljeryd, P. Ekstrand, F. Henn, and K. Kjörling, "Source coding enhancement using spectral band replication," EP0940015B1, 2004.

[6] L. Liljeryd, P. Ekstrand, F. Henn, and K. Kjörling, "Improved spectral translation," EP1285436B1, 2003.

[7] L. Liljeryd, P. Ekstrand, F. Henn, and K. Kjörling, "Efficient spectral envelope coding using variable time/frequence resolution," EP1216474B1, 2004.

[8] L. Liljeryd, P. Ekstrand, F. Henn, and K. Kjörling, "Enhancing perceptual performance of SBR and related HFR coding methods by adaptive noise-floor addition and noise substitution limiting," EP1157374B1, 2004.

[9] K. Kjörling, P. Ekstrand, F. Henn, and L. Villemoes, "Enhancing perceptual performance of high frequency reconstruction coding methods by adaptive filtering," EP1342230B1, 2004.

[10] A. Gröschel, M. Schug, M. Beer, and F. Henn, "Enhancing audio coding efficiency of MPEG Layer-2 with spectral band replication for digital radio (DAB) in a backwards compatible way," in *Proceedings of the 114th Convention of the Audio Engineering Society (AES '03)*, Amsterdam, The Netherlands, March 2003, paper number 5850.

[11] T. Ziegler, A. Ehret, P. Ekstrand, and M. Lutzky, "Enhancing mp3 with SBR: features and capabilities of the new mp3PRO algorithm," in *Proceedings of the 112th Convention of the Audio Engineering Society (AES '02)*, Munich, Germany, May 2002, paper number 5560.

[12] European Telecommunications Standards Institute, "Digital Radio Mondiale (DRM); System Specification," ETSI ES 201 980 V2.2.1, October 2005.

[13] European Telecommunications Standards Institute, "Universal Mobile Telecommunications System (UMTS); General audio codec audio processing functions; Enhanced aacPlus general audio codec; Encoder specification; Spectral Band Replication (SBR) part (3GPP TS 26.404 version 6.0.0 Release 6)," ETSI TS 126 404 V6.0.0, September 2004.

[14] P. Ekstrand, "Bandwidth extension of audio signals by spectral band replication," in *Proceedings of the 1st IEEE Benelux Workshop on Model Based Processing and Coding of Audio (MPCA '02)*, Leuven, Belgium, November 2002.

[15] 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.

[16] A. Ehret, K. Kjörling, J. Rödén, H. Purnhagen, and H. Hörich, "aacPlus, only a low-bitrate codec?" in *Proceedings of the 117th Convention of the Audio Engineering Society (AES '04)*, San Francisco, Calif, USA, October 2004, paper number 6199.

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

[18] ITU-T Recommend. P800, "Methods for subjective determination of transmission quality," 1996.

[19] ISO/IEC JTC1/SC29/WG11, "Report on the verification tests of MPEG-4 High Efficiency AAC," ISO/IEC JTC1/SC29/WG11 N6009, October 2003.

[20] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proceedings of the IEEE*, vol. 88, no. 4, pp. 451–512, 2000.

[21] A. S. Spanias, "Speech coding: a tutorial review," *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1541–1582, 1994.

[22] M. Ali, *Adaptive signal representation with application in audio coding*, Ph. D. thesis, University of Minnesota, Minneapolis, Minn, USA, 1996.

[23] P. Masri, *Computer modelling of sound for transformation and synthesis of musical signals*, Ph. D. thesis, University of Bristol, Bristol, UK, 1996.

[24] M. M. Goodwin, *Adaptive signal models: theory, algorithms, and audio applications*, Ph.D. thesis, University of California, Berkeley, Calif, USA, 1997.

[25] S. N. Levine, *Audio representation for data compression and compressed domain processing*, Ph. D. thesis, Stanford University, Stanford, Calif, USA, 1999.

[26] D. V. Anderson, *Audio signal enhancement using multi-resolution sinusoidal modeling*, Ph. D. thesis, Georgia Institute of Technology, Atlanta, Ga, USA, 1999.

[27] F. P. Myburg, *Design of a scalable parametric audio coder*, Ph. D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2004.

[28] R. Vafin, *Towards flexible audio coding*, Ph. D. thesis, KTH (Royal Institute of Technology), Stockholm, Sweden, 2004.

[29] M. G. Christensen, *Estimation and modeling problems in parametric audio coding*, Ph. D. thesis, Aalborg University, Aalborg, Denmark, 2005.

[30] H. Purnhagen and N. Meine, "HILN-the MPEG-4 parametric audio coding tools," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS '00)*, vol. 3, pp. 201–204, Geneva, Switzerland, May 2000.

[31] ISO/IEC, "Coding of audio-visual objects—part 3: audio, AMENDMENT 1: audio extensions (MPEG-4 Audio Version 2)," ISO/IEC Int. Std. 14496-3:1999/Amd.1:2000, 2000.

[32] P. Hedelin, "A tone-oriented voice-excited vocoder," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '81)*, vol. 1, pp. 205–208, Atlanta, Ga, USA, March 1981.

[33] L. Almeida and J. Tribolet, "Nonstationary spectral modeling of voiced speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 31, no. 3, pp. 664–678, 1983.

[34] R. McAulay and T. Quartieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[35] X. Serra and J. Smith III, "Spectral modeling synthesis. A sound analysis/synthesis system based on a deterministic plus stochastic decomposition," *Computer Music Journal*, vol. 14, no. 4, pp. 12–24, 1990.

[36] ISO/IEC, "Coding of audio-visual objects. Part3: audio, AMENDMENT 2: parametric coding of high quality audio," ISO/IEC Int. Std. 14496-3:2001/Amd2:2004, July 2004.

[37] A. C. den Brinker, E. G. P. Schuijers, and A. W. J. Oomen, "Parametric coding for high-quality audio," in *Proceedings of the 112th Convention of the Audio Engineering Society (AES '02)*, Munich, Germany, May 2002, paper number 5554.

[38] E. G. P. Schuijers, A. W. J. Oomen, A. C. den Brinker, and A. J. Gerrits, "Advances in parametric coding for high-quality audio," in *Proceedings of the 1st IEEE Benelux Workshop on Model Based Processing and Coding of Audio (MPCA '02)*, pp. 73–79, Leuven, Belgium, November 2002.

[39] E. G. P. Schuijers, A. C. den Brinker, A. W. J. Oomen, and D. J. Breebaart, "Advances in parametric coding for high-quality audio," in *Proceedings of the 114th Convention of the Audio Engineering Society (AES '03)*, Amsterdam, The Netherlands, March 2003, paper number 5852.

[40] A. C. den Brinker and F. Riera-Palou, "Pure linear prediction," in *Proceedings of the 115th Convention of the Audio Engineering Society (AES '03)*, New York, NY, USA, October 2003, paper number 5924.

[41] S. N. Levine and J. O. Smith III, "A sines+transients+noise audio representation for data compression and time/pitch scale modifications," in *Proceedings of the 105th Convention of the Audio Engineering Society (AES '98)*, San Francisco, Calif, USA, September 1998, paper number 4781.

[42] J. Kliewer and A. Mertens, "Audio subband coding with improved representation of transient signal segments," in *Proceedings of the 9th European Signal Processing Conference (EUSIPCO '98)*, S. Theodoridis, I. Pitas, A. Stouraitis, and N. Kalouptsidis, Eds., pp. 2345–2348, Rhodos, Greece, September 1998.

[43] E. B. George and M. J. T. Smith, "A new speech coding model based on a least-squares sinusoidal representation," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '87)*, pp. 1641–1644, Dallas, Tex, USA, April 1987.

[44] M. Goodwin, "Matching pursuit with damped sinusoids," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, vol. 3, pp. 2037–2040, Munich, Germany, April 1997.

[45] A. C. den Brinker, A. J. Gerrits, and R. J. Sluijter, "Phase transmission in sinusoidal audio and speech coding," in *Proceedings of the 115th Convention of the Audio Engineering Society (AES '03)*, New York, NY, USA, October 2003, paper number 5983.

[46] H. Purnhagen, N. Meine, and B. Edler, "Sinusoidal coding using loudness-based selection," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '02)*, vol. 2, pp. 1817–1820, Orlando, Fla, USA, May 2002.

[47] J. O. Smith III and J. S. Abel, "Bark and ERB bilinear transforms," *IEEE Transactions on Speech and Audio Processing*, vol. 7, no. 6, pp. 697–708, 1999.

[48] J. Herre and J. D. Johnston, "Enhancing the performance of perceptual audio coders by using temporal noise shaping," in *Proceedings of the 101st Convention of the Audio Engineering Society (AES '96)*, Los Angeles, Calif, USA, November 1996, paper number 4384.

[49] M. Athineos and D. P. W. Ellis, "Autoregressive modeling of temporal envelopes," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5237–5245, 2007.

[50] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, San Diego, Calif, USA, 1997.

[51] J. D. Johnston and A. J. Ferreira, "Sum-difference stereo transform coding," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '92)*, pp. 569–572, San Francisco, Calif, USA, March 1992.

[52] J. Herre, K. Brandenburg, and D. Lederer, "Intensity stereo coding," in *Proceedings of the 96th Convention of the Audio Engineering Society (AES '94)*, Amsterdam, The Netherlands, February-March 1994, paper number 3799.

[53] J. Breebaart, S. van de Par, A. Kohlrausch, and E. Schuijers, "Parametric coding of stereo audio," *EURASIP Journal on Applied Signal Processing*, vol. 2005, no. 9, pp. 1305–1322, 2005.

[54] J. Breebaart and C. Faller, *Spatial Audio Processing: MPEG Surround and Other Applications*, John Wiley & Sons, Chichester, 2007.

[55] E. Schuijers, J. Breebaart, H. Purnhagen, and J. Engdegård, "Low complexity parametric stereo coding," in *Proceedings of the 116th Convention of the Audio Engineering Society (AES '04)*, Berlin, Germany, May 2004, paper number 6073.

[56] J. Engdegård, H. Purnhagen, J. Rödén, and L. Liljeryd, "Synthetic ambience in parametric stereo coding," in *Proceedings of the 116th Convention of the Audio Engineering Society (AES '04)*, Berlin, Germany, May 2004.

[57] H. Purnhagen, "Low complexity parametric stereo coding in MPEG-4," in *Proceedings of Digital Audio Effects Workshop (DAFX)*, Naples, Italy, October 2004.

[58] ISO/IEC JTC1/SC29/WG11, "Report on the verification test of MPEG-4 parametric coding for high-quality audio," ISO/IEC JTC1/SC29/WG11 N6675, 2004.

[59] 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, December 2003.

[60] ISO/IEC, "Coding of audio-visual objects—part 3: audio, AMENDMENT 2: audio lossless coding (ALS), new audio profiles and BSAC extensions," ISO/IEC Int. Std. 14496-3:2005/Amd.2:2006, 2006.

[61] 3rd Generation Partnership Project, "3GPP TS 26.401 V6.2.0, 3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; General audio codec audio processing functions; Enhanced aacPlus general audio codec; General description," March 2005.

EURASIP Journal on Wireless Communications and Networking

Special Issue on High-Throughput Wireless Baseband Processing

Call for Papers

Wireless communications is a fast-paced area, where many standards, protocols, and services are introduced each year. Implementation of every new standard becomes challenging especially when more and more higher data rates up to several gigabits/second are required. On the other hand, the power budget is not increasing in the same pace. The presence of all those different modes as well as high throughput requirements brought the need for designing almost-all-digital radios, which benefit from technology scaling. Those goals can only be achieved by efficient algorithms, models, and methods for the design of high-throughput and low-power systems for baseband processing. This special issue will report the recent advances of very high throughput and low-power systems for wireless baseband processing. Areas of interest include, but are not limited to:

- Modeling of quality-of-service, reliability, and performance in high-throughput wireless systems
- Power-aware and/or low-cost algorithms and architecture optimizations for multistandard baseband processing
- Baseband compensation techniques for RF/analog circuit impairments
- High-throughput baseband processing for software-defined and cognitive radios
- Applications to WirelessHD, IEEE 802.15.3c, MIMO systems, UWB, WiMAX, and LTE systems

Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://www.hindawi.com/journals/wcn/guidelines.html. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/ according to the following timetable:

Manuscript Due	October 1, 2009
First Round of Reviews	January 1, 2010
Publication Date	April 1, 2010

Lead Guest Editor

Taskin Kocak, University of Bristol, Bristol, UK; t.kocak@bristol.ac.uk

Guest Editors

Mustafa Badaroglu, ON Semiconductor Vilvoorde, Belgium; mustafa.badaroglu@onsemi.com

Dake Liu, Linkoping University, Linkoping, Sweden; dake@isy.liu.se

Liesbet Van der Perre, IMEC, Leuven, Belgium; vdperre@imec.be

EURASIP Journal on Wireless Communications and Networking

Special Issue on

Simulators and Experimental Testbeds Design and Development for Wireless Networks

Call for Papers

In the context of wireless networking, performance evaluation of protocols and distributed applications is generally conducted through simulation or experimentation campaigns. An efficient and accurate simulation of wireless networks raises various issues which generally need to be addressed from several research domains simultaneously. As examples, we can consider the wireless physical layer modeling and simulation, the support of large-scale networks, the simulation of complex RF systems such as MIMO ones, the emulation of wireless nodes or the interconnection of simulators, experimental testbeds, and so forth.

The aim of this Special Issue is to bring together academic and industry researchers and practitioners from both the wireless networking and the simulation communities to discuss current and future trends in simulation or experimentation techniques, models, and practices for the future communication system and to foster interdisciplinary collaborative research in this area. The guest editors seek high-quality papers on aspects of wireless network simulation, and value both theoretical and practical research contributions. Topics of interest include, but are not limited to:

- Radio medium modeling and cross-layer simulation
- Scalability, large-scale networks support
- Validation of simulators and simulation results
- Simulators benchmarking and comparisons
- Fluid-flow simulation for assessing QoS in large-scale networks
- Support of new emerging technologies (WiMax, 3.5G, Wireless Mesh Networks, 802.11x, etc.) in simulators
- Support of advanced RF systems (Multi-carrier schemes, MIMO, smart-antenna) in simulators
- Wireless node simulation or emulation
- Interoperability of simulators, emulators, and experiments
- Support of distributed physical layer schemes (distributed signal processing; cooperative schemes)
- Distributed simulation, and scalability of simulators
- Implementation of simulators
- Experimental testbeds for wireless networks
- Methodology for protocol and distributed application performance evaluation
- SDR techniques, cognitive radio approaches, dynamic spectrum access testbeds, and simulators as well as modeling
- Simulation and testbeds for cooperative communication protocols

Before submission, authors should carefully read over the journal's Author Guidelines, which are located at http://www.hindawi.com/journals/wcn/guidelines.html. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/, according to the following timetable:

Manuscript Due	June 1, 2009
First Round of Reviews	September 1, 2009
Publication Date	December 1, 2009

Lead Guest Editor

Faouzi Bader, Centre Tecnológico de Telecomunicaciones de Cataluya (CTTC), PMT, Spain; faouzi.bader@cttc.es

Guest Editors

Guillame Chelius, Institut National de Recherche en Informatique et en Automatique (INRA), France; guillaume.chelius@inria.fr

Christian Ibars, Centre Tecnológico de Telecomunicaciones de Cataluya (CTTC), PMT, Spain; cibars@cttc.cat

Mohamed Ibnkahla, Department of Electrical and Computer Engineering, Queen's University, Canada; ibnkahla@post.queensu.ca

Nikos Passas, Department of Informatics and Telecommunications, University of Athens, Greece; passas@di.uoa.gr

Arnd-Ragnar Rhiemeier, Defence Electronics, EADS Deutschland GmbH, Germany; arnd-ragnar.rhiemeier@eads.com

EURASIP Journal on Wireless Communications and Networking

Special Issue on

Interference Management in Wireless Communication Systems: Theory and Applications

Call for Papers

Interference is a fundamental nature of wireless communication systems, in which multiple transmissions often take place simultaneously over a common communication medium. In recent years, there has been a rapidly growing interest in developing reliable and spectral efficient wireless communication systems. One primary challenge in such a development is how to deal with the interference, which may substantially limit the reliability and the throughput of a wireless communication system. In most existing wireless communication systems, interference is dealt with by coordinating users to orthogonalize their transmissions in time or frequency, or by increasing transmission power and treating each other's interference as noise. Over the past twenty years, a number of sophisticated receiver designs, for example, multiuser detection, have been proposed for interference suppression under various settings. Recently, the paradigm has shifted to focus on how to intelligently exploit the knowledge and/or the structure of interference to achieve improved reliability and throughput of wireless communication systems.

This special issue aims to bring together state-of-the-art research contributions and practical implementations that effectively manage interference in wireless communication systems. Original contributions in all areas related to interference management for wireless communication systems are solicited for this special issue. We are particularly interested in manuscripts that report the latest development on interference channels or cognitive radio channels from the perspectives of information theory, signal processing, and coding theory. Topics of interest include, but are not limited to:

- Information theoretic study of interference channels or cognitive radio channels
- Game theoretical approach to interference management in wireless networks
- Cooperative wireless communication systems
- Relaying in interference networks
- Advanced coding schemes for interference/cognitive radio channels
- Interference channels with confidentiality requirement
- Femtocell networks
- Signal processing algorithms for interference mitigation
- Receiver designs for interference channels or cognitive radio channels
- MIMO interference channels or MIMO cognitive radio channels
- Base station cooperation for interference mitigation
- Network coding for interference channels or cognitive radio channels
- Resource allocation for interference management

Before submission authors should carefully read over the journal's Author Guidelines, which are located at http://www.hindawi.com/journals/wcn/guidelines.html. Prospective authors should submit an electronic copy of their complete manuscript through the journal Manuscript Tracking System at http://mts.hindawi.com/ according to the following timetable:

Manuscript Due	November 1, 2009
First Round of Reviews	February 1, 2010
Publication Date	June 1, 2010

Lead Guest Editor

Yan Xin, NEC Laboratories America, Inc., Princeton, NJ 08540, USA; yanxin@nec-labs.com

Guest Editors

Xiaodong Wang, Electrical Engineering Department, Columbia University, New York, NY 10027, USA; wangx@ee.columbia.edu

Geert Leus, Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands; g.j.t.leus@tudelft.nl

Guosen Yue, NEC Laboratories America, Inc., Princeton, NJ 08540, USA; yueg@nec-labs.com

Jinhua Jiang, Department of Electrical Engineering, Stanford University Stanford, CA 94305, USA; jhjiang@stanford.edu