BANDWIDTH EXTENSION OF AUDIO SIGNALS BY SPECTRAL BAND REPLICATION

Per Ekstrand

Coding Technologies
Döbelnsgatan 64, 113 52 Stockholm, Sweden
pe@codingtechnologies.com

ABSTRACT

Spectral Band Replication (SBR) is a new audio coding tool that significantly improves the coding gain of perceptual coders and speech coders. Currently, there are three different audio coders that have shown a vast improvement by the combination with SBR: MPEG-AAC, MPEG-Layer II and MPEG-Layer III (mp3), all three being parts of the open ISO-MPEG standard. The combination of AAC and SBR will be used in the standardized Digital Radio Mondiale (DRM) system, and SBR is currently also being standardized within MPEG-4. SBR is a so-called bandwidth extension technique, where a major part of a signal's bandwidth is reconstructed from the lowband on the receiving side. It is developed and marketed by Coding Technologies, an international company in the audio coding field. This paper will focus on the technical details of SBR and in particular on the filter bank, which is the basis of the SBR process.

1. INTRODUCTION

Spectral Band Replication (SBR) [1] is a bandwidth extension method originally invented and developed in Sweden by Coding Technologies (CT). Having grown from a four-employee company in 1997, CT is today one of the leading audio coding companies, with offices in Sweden, Germany and USA.

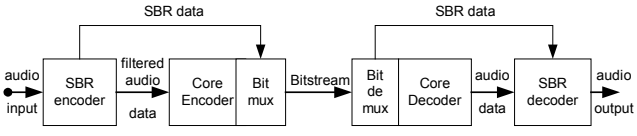

Figure 1: SBR is a preprocess to the core encoder and a postprocess to the core decoder.

SBR is not a self-contained audio coder - it is an add-on to a traditional audio or speech coder, hereinafter referred to as a *core* coder. SBR acts as a preprocess to the core encoder, and as a postprocess to the core decoder, Fig. 1. On the encoder side, SBR extracts vital guidance information to ensure optimal operation of the SBR process on the decoder side. This information, referred to as SBR data, has a very moderate data rate, typically a small fraction of the data rate of the combined system. The core coder is responsible for coding the *lowband*, a limited bandwidth of the original audio signal up to a certain *cutoff frequency*. All frequencies above the cutoff frequency are, by the SBR process, reconstructed from the lowband in a perceptually accurate way to form the *highband*. The combination of the highband with the lowband results in a full bandwidth decoded audio signal. Generally, the core coder operates at half the sampling rate of SBR. This results in higher frequency resolution for the core coder filter bank and consequently improved means for utilization of the auditory masking effects.

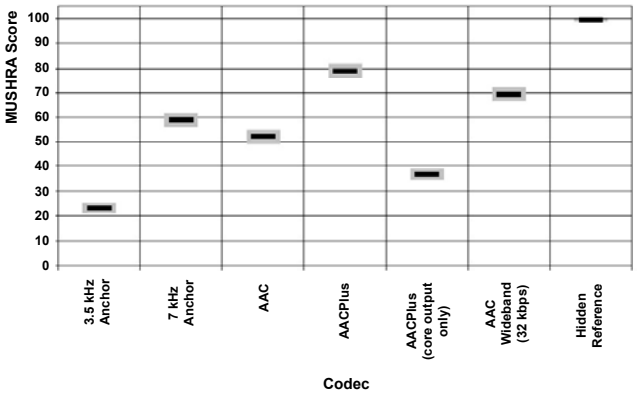

Figure 2: DRM test results for 24 kbps mono items. BBC test site. Note that the bitrate of the AAC Wideband coder was 32 kbps. In the MUSHRA test, a score above 80 is labeled excellent.

The general idea behind SBR is that there exists a strong correlation between the highband characteristics of a signal with the lowband characteristics of the same signal. A signal with a strong harmonic series reaching up to the cutoff frequency is naturally assumed to consist of the same harmonic series in the highband, although maybe not as pronounced as in the lowband. A noise-like signal in the lowband is in the same manner assumed to keep its noisiness in the highband. This rule generally gives the best estimation for the highband characteristics. Of course, there are signals that deviate from this model. Luckily, SBR has methods to handle those awkward situations. Inverse filtering, adaptive noise addition and sinusoidal regeneration are tools to improve signals that have less correlation between the characteristics of the lowband and the highband.

The SBR system is composed of several modules, each having a specific purpose. This and the fact that SBR recreates the highband from the lowband with the help of low rate guidance information makes SBR closely related to parametric methods. However, there is no need to track sinusoidals or take care of birth and death-processes for partials as is usually the case in parametric coding. In SBR, the short-term synchronization of the highband with the lowband, i.e. the time alignment, is close to optimal due to the nature of the highband generation. A transient in the lowband will translate almost perfectly to the highband. A sinusoidal present in the highband will persist in time as long as its corresponding partial exists in the lowband.

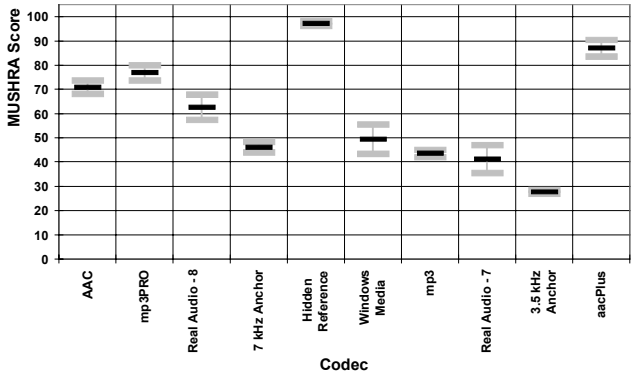

Figure 3: EBU test results for 48 kbps stereo items. Mean values and 95 % confidence intervals. IRT test site.

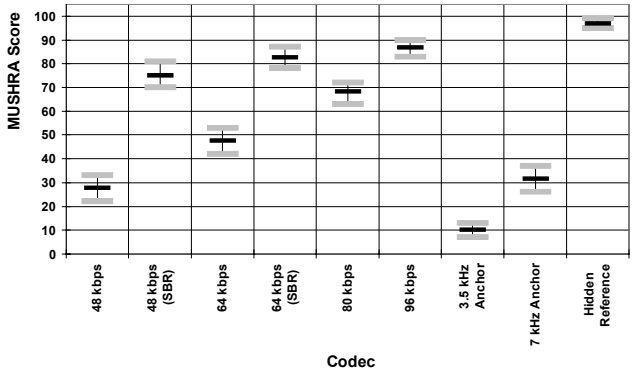

Figure 4: IRT mono test results for various bitrates of MPEG-Layer I/II.

SBR significantly improves the coding efficiency of perceptual coders or speech coders. Coding Technologies has integrated SBR into three different waveform coders: MPEG-AAC, MPEG-Layer II and MPEG-Layer III (mp3) – all being parts of the open ISO-MPEG standard [2] [3]. Several independent listening tests have shown a coding gain by more than 30 % using the SBR-enhanced coders. The combination of SBR and AAC, aacPlus, is currently the world's most efficient audio coder [4]. The aacPlus coder is used by XM Satellite Radio in their satellite-based digital broadcasting system [5]. XM Radio has more than 200.000 subscribers for their radio services and the number is rapidly growing. The Digital Radio Mondiale (DRM) consortium has chosen aacPlus as its audio source coder [6]. DRM is a digital broadcasting standard for frequencies below 30 MHz (long, medium and short-wave). The DRM specification was finalized in January 2001 as an ETSI standard [7]. The SBR-enhanced version of mp3, mp3PRO, is marketed by Thomson Multimedia, and has found its way into several commercial products, both software and hardware based [8]. Both aacPlus and mp3PRO have thoroughly been evaluated in several listening tests. Fig. 2 shows the result of a test conducted for DRM and Fig. 3 shows result from the European Broadcasting Union (EBU). MPEG-Layer II + SBR has been evaluated in listening tests by IRT, Fig. 4. MPEG-Layer II is the audio source coding method used in the DAB standard [9]. SBR itself was submitted as a proposal to MPEG in January 2001, and got accepted as Reference Model 0 shortly afterwards. Today, it has reached Final Preliminary Draft Amendment (FPDAM) status [10].

By transmitting the SBR data as ancillary data in the core coder bitstream, all SBR-enhanced coders have the obvious advantage of being backward and forward compatible to the core coder standard. This permits the introduction of SBR to existing systems, already in operation using the plain core coder, thus enabling a smooth transition from traditional audio coding to the more efficient SBR-enhanced version. Old decoders will of course not benefit from the new technique but will still be able to output the bandlimited signal from the core decoder.

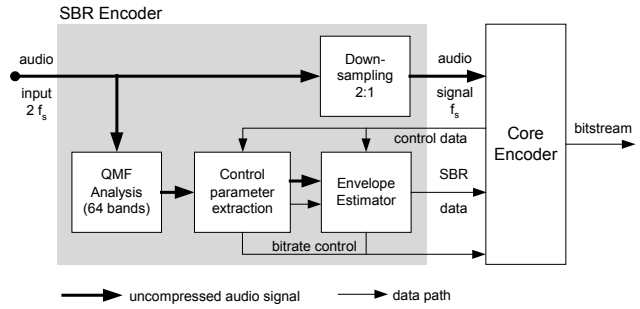

Figure 5: Encoder block diagram. The SBR encoder interacts closely with the core encoder.

2. THE SBR ENCODER

Although SBR is mainly a postprocess, vital parameters are extracted in the encoder to assure optimal operation of the SBR decoder. The basic layout of the SBR encoder is depicted in the block diagram of Fig. 5. The input signal is initially fed to a downsampler, which supplies the core encoder with a time domain signal having half the sampling frequency of the input signal. The input signal is in parallel fed to a 64-channel analysis QMF bank. The outputs from the filter bank are complex-valued subband signals. The subband signals are fed to an envelope estimator and various detectors. The outputs from the detectors and the envelope estimator are assembled into the SBR data stream. The data is subsequently coded using entropy coding and, in the case of multichannel signals, also channel-redundancy

coding. The coded SBR data and a bitrate control signal are then supplied to the core encoder. The SBR encoder interacts closely with the core encoder. Information is exchanged between the systems in order to, for example, determine the optimal cutoff frequency between the core coder and the SBR band. The core coder finally multiplexes the SBR data stream into the combined bitstream.

3. THE SBR DECODER

The block diagram of Fig. 6 illustrates the layout of the SBR enhanced decoder. The received bitstream is divided into two parts: the core coder bitstream and the SBR data stream. The core bitstream is decoded by the core decoder, and the output audio signal, typically of lowpass character, is forwarded to the SBR decoder together with the SBR data stream. The core audio signal, 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. The output from the filter bank, i.e. the subband signals, are complex-valued and thus oversampled by a factor of two compared to a regular QMF bank. The technical details of the QMF bank are covered in the next section.

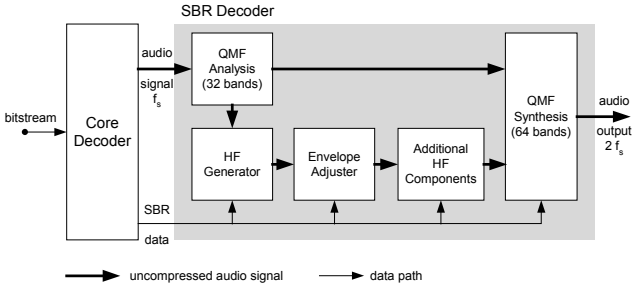

Figure 6: Basic decoder structure. SBR acts as a post-process to the core decoder.

The complex-valued subband signals obtained from the filter bank are processed in the high frequency generation unit to obtain a set of highband subband signals. The generation is performed by selecting lowband subband signals, according to specific rules, for the conversion to highband signals. As mentioned in the previous chapter, the correlation of the characteristics between the lowband and the highband varies for different audio signals. The tonality is for example usually more pronounced in the lowband than in the highband. Therefore, inverse filtering is applied to the generated subband signals. The filtering is accomplished through inband filtering of the complex-valued signals using adaptive low-order complex-valued FIR filters. The filter coefficients are determined through an analysis of the lowband in combination with control signals extracted from the SBR data stream. The generated highband signals are subsequently fed to the envelope adjusting unit.

The most important, and also the largest part of the SBR data stream is the short-term spectral envelope representation of the highband. This envelope representation is used to adjust the energy of the newly generated highband subband signals. The envelope adjusting unit first performs an energy estimate of the highband 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 highband subband samples comprised in the respective segments are adjusted.

According to the above, the estimated envelope samples are obtained by averaging of subband sample energies within segments. The time and frequency borders of these segments, expressed in subband sample indices, are determined by examination of the input signal properties on the encoder side, and subsequently signaled to the decoder. Generally, longer segments of higher frequency resolution are used during quasi-stationary passages, and smaller segments of lower frequency resolution are used for transient-like passages. 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 filter bank design. The filter bank is hence constructed to provide a resolution in both time and frequency that is considered adequate for the adjustment of the envelope, at any time instant. The filter bank resolution is not adaptive, as is usually the case for filter banks in perceptual waveform coders. Instead, the energy estimates are grouped adaptively within a filter bank of fixed size. This grouping of subband samples can be carried out instantaneously, without the requirement for the filter bank to switch from one state, i.e. time/frequency resolution, to another. The maximum resolution possible is thus the time and frequency support given by one subband sample or one filter bank channel.

Sinusoidals present in the original highband signal that has no corresponding sinusoidal in the generated highband is synthesized using the sinusoidal regenerator. Subsequently, random white noise is added to the highband signals to compensate for diverting tonal to noise ratios of the highband and lowband. Both the sinusoidal regeneration and the adaptive noise addition are controlled by signals extracted from the SBR data stream.

The generated highband signals and the delayed lowband signals are finally supplied to the 64-channel synthesis filter bank, which operates at the output sampling frequency. The filter bank generates a real-valued full bandwidth output signal having twice the sampling frequency of the core coder signal.

4. THE COMPLEX QMF BANK

Impairments emerging from modifications of real-valued subband signals can be significantly reduced by extending a cosine modulated filter bank with an imaginary sine modulated part, forming a complex-exponential modulated filter bank. The sine extension eliminates the main alias terms present in the cosine modulated filter bank. 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.

A method for optimization of the complex modulated filter bank, referred to as alias term minimization (ATM), has been developed to minimize the remaining alias components when using the filter bank as an equalizer.

4.1. Cosine Modulated Filter Banks

In a cosine modulated filter bank [11] the analysis filters $h_k(n)$ are cosine modulated versions of a symmetric low-pass prototype filter $p_0(n)$ as

$$h_k(n) = p_0(n)\cos\left\{\frac{\pi}{2M}(2k+1)(n-\frac{N}{2}-\frac{M}{2})\right\} \qquad (1)$$

where $k = 0 \ldots M\text{-}1$, M is the number of channels and $n = 0 \ldots N$, where N is the prototype filter order.

Following the same notation, the synthesis filters are given by

$$f_k(n) = p_0(n)\cos\left\{\frac{\pi}{2M}(2k+1)(n-\frac{N}{2}+\frac{M}{2})\right\} \qquad (2)$$

The analysis filter bank produces real-valued subband samples for real-valued input signals. The subband samples are downsampled a factor M, making the system critically sampled. Depending on the choice of the prototype filter, the filter bank may constitute a near perfect reconstruction system, a so-called pseudo QMF bank, or a perfect reconstruction (PR) system. An example of a pseudo QMF bank is the filter bank used in MPEG-Layer I/II and an example of a PR system is the modulated lapped transform (MLT) [12]. One inherent property of the cosine modulation is that every filter has two passbands: one in the positive frequency range and one corresponding passband in the negative frequency range. It is easily shown that the main alias terms emerge from overlap in frequency between either the filters negative passband with frequency modulated versions of the positive passband, or vice versa. The cosine modulated banks offer very effective implementations and are often used in natural audio codecs [13]. However, any attempt to alter the subband samples or spectral coefficients, e.g. by applying an equalizing gain curve, renders severe aliasing artifacts in the output signal.

4.2. Complex-Exponential Modulated Filter Banks

Extending the cosine modulation to complex-exponential modulation yields the analysis filters $h_k(n)$ as

$$h_k(n) = p_0(n)\exp\left\{i\ \frac{\pi}{2M}(2k+1)(n-\frac{N}{2}-\frac{M}{2})\right\} \qquad (3)$$

using the same notation as before. This can be viewed as adding an imaginary part to the real-valued filter bank, where the imaginary part consists of sine modulated versions of the same prototype filter. Considering a real-valued input signal, the output from the filter bank can be interpreted as a set of subband signals, where the real and the imaginary parts are Hilbert transforms of each other. The resulting subbands are thus the analytic signals of the real-valued output obtained from the cosine modulated filter bank. Due to the complex-valued representation, the subband signals are oversampled by a factor two compared to the real-valued versions.

The synthesis filters are extended in the same way as

$$f_k(n) = p_0(n)\exp\left\{i\ \frac{\pi}{2M}(2k+1)(n-\frac{N}{2}+\frac{M}{2})\right\} \qquad (4)$$

Eq.(3) and (4) implies that the output from the synthesis bank is complex-valued. Using matrix notation, where $\mathbf{C}_a$ is a matrix with analysis filters from Eq.(1), and $\mathbf{S}_a$ is a matrix with filters as

$$h_k(n) = p_0(n)\sin\left\{\frac{\pi}{2M}(2k+1)(n-\frac{N}{2}-\frac{M}{2})\right\} \qquad (5)$$

the filters of Eq.(3) is obtained as $\mathbf{C}_a + \mathrm{j}\ \mathbf{S}_a$. In these matrices, k is the row index and n is the column index. Analogously, the matrix $\mathbf{C}_s$ has synthesis filters from Eq.(2), and $\mathbf{S}_s$ is a matrix with filters as

$$f_k(n) = p_0(n)\sin\left\{\frac{\pi}{2M}(2k+1)(n-\frac{N}{2}+\frac{M}{2})\right\} \qquad (6)$$

Eq.(4) can thus be written $\mathbf{C}_s + \mathrm{j}\ \mathbf{S}_s$, where k is the column index and n is the row index. Denoting the input signal $\mathbf{x}$, the output signal $\mathbf{y}$ can be found from

$$\begin{aligned}\mathbf{y} &= (\mathbf{C}_s + \mathrm{j}\ \mathbf{S}_s)\,(\mathbf{C}_a + \mathrm{j}\ \mathbf{S}_a)\,\mathbf{x} = \\ &= (\mathbf{C}_s\mathbf{C}_a - \mathbf{S}_s\mathbf{S}_a)\,\mathbf{x} + \mathrm{j}\,(\mathbf{C}_s\mathbf{S}_a + \mathbf{S}_s\mathbf{C}_a)\,\mathbf{x}\end{aligned} \qquad (7)$$

As seen from Eq.(7), the real part consists of two terms; the output from the ordinary cosine modulated filter bank and an output from a sine modulated filter bank. It is easily verified that if a cosine modulated filter bank has the PR property, then its sine modulated version, with a change of sign, constitutes a PR system as well. Thus, by taking the real part of the output, the complex-exponential modulated system offers the same reconstruction accuracy as the corresponding cosine modulated version.

The complex-exponential modulated system can be extended to handle also complex-valued input signals. By extending the number of channels to $2M$, i.e. adding the filters for the negative frequencies, and keeping the imaginary part of the output signal, a pseudo QMF or a PR system for complex-valued signals is obtained.

The complex-exponential modulated filter bank has only one passband for every channel and is thus free from main alias terms. The absence of main alias terms makes the aliasing cancellation constraint from the cosine (or sine) modulated filter bank obsolete in the complex-exponential modulated version. Both the analysis and synthesis filters can thus be written

$$h_k(n) = f_k(n) = p_0(n)\exp\left\{i\ \frac{\pi}{2M}(2k+1)(n-\frac{N}{2})\right\} \qquad (8)$$

As before, $k = 0 \ldots M\text{-}1$, where M is the number of channels and $n = 0 \ldots N$, where N is the prototype filter order.

Due to the absence of main alias terms, the resulting aliasing is dependent only on the suppression of the alias terms emanating from overlap between filters and their modulated versions. It is thus of great importance to design the prototype filter for maximum suppression of these terms. This is preferably accomplished by optimizing the prototype filter using standard nonlinear optimization algorithms, for example the Downhill Simplex Method [14]. In these algorithms, the idea is to minimize an objective function. The specification of this function is crucial for the success of the optimization. The alias term minimization method significantly improves the performance of the filter banks as we will see below. The details of the optimization, however, are not covered by this paper.

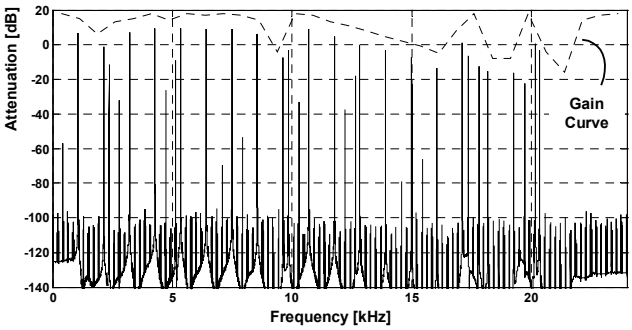

Figure 7: MPEG-Layer I/II filter bank operated as an equalizer. The dotted trace shows the equalizing curve (in dB). The solid trace shows an equalized harmonic series (1070 Hz fundamental). The output is seriously contaminated by aliasing.

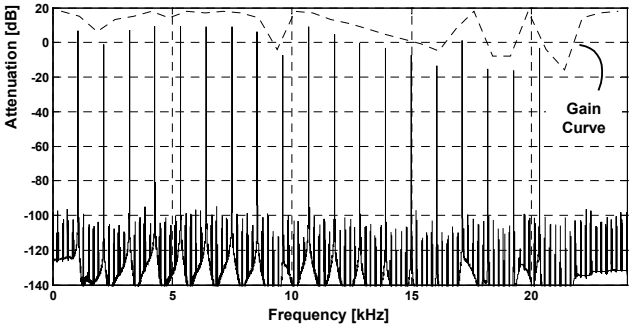

Figure 8: Complex-exponential modulated version of the MPEG-Layer I/II filter bank operated as an equalizer. The output has remaining alias components at approximately -97 dB.

4.3. Modification of Subband Signals

The constraint for PR filter banks [11] imposes large limitations for a filter bank used in an equalization system. The prototype filter design is then restricted to follow the PR property, which results in less degree of freedom during the optimization. A pseudo QMF system, on the other hand, can always be designed for adequate reconstruction accuracy, since all practical implementations have limited numerical resolution.

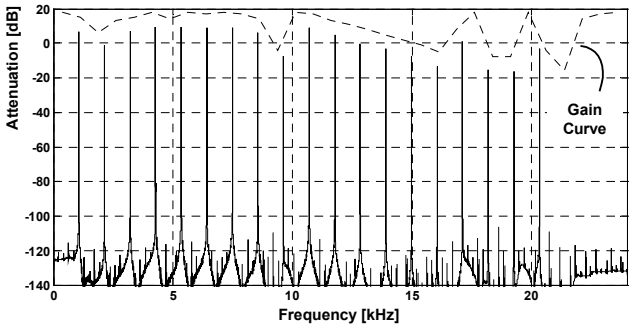

Figure 9: Complex-exponential modulated filter bank with alias term minimized prototype filter operated as an equalizer. The remaining alias components are suppressed additionally 21 dB compared to Fig .8.

We will compare three filter banks used as equalizers. We commence by applying an equalization curve to the cosine modulated filter bank used in MPEG-Layer I/II. The filter bank has 32 channels and a prototype filter of order 512. The equalizing gain curve is showed as the dotted trace in Fig. 7. The solid trace shows the output from the filter bank. The input signal is a harmonic series with a flat envelope and a fundamental frequency of 1070 Hz sampled at 48 kHz. The aliasing in the output signal is evident and results in a total aliasing attenuation of only 15 dB compared to a full rate, i.e. non-downsampled, filter bank. Moving to complex-exponential modulation of the same filter bank gives an aliasing attenuation of 97 dB and the output shown in Fig. 8. It is obvious that a substantial improvement is achieved by the complex-valued representation. Fig. 9 shows the same equalized signal, but this time processed with a complex modulated filter bank with an alias term minimized prototype filter of the same order. The total rejection of aliasing is 118 dB – an improvement compared to the already very efficient MPEG-type filter by 21 dB. The passband flatness and the aliasing attenuation are compared in Table 1, which also shows the total error rejection of the filter banks. Note that the values in Table 1 consider non-equalized filter banks.

Table 1: Performance of non-equalized 32-channel filter banks.

	Cosine modulated	Complex modulated	ATM Complex modulated
Passband flatness	84.6 dB	84.6 dB	84.9 dB
Aliasing rejection	97.4 dB	97.4 dB	133.3 dB
Total error rejection	84.4 dB	84.4 dB	84.9 dB

5. CONCLUSION

Spectral Band Replication is a new technology that substantially improves the performance of existing and future audio coders. The combination of AAC and SBR, aacPlus, is the most efficient audio coder today, improving the already powerful AAC coder in coding efficiency by more than 30 %. The foundation of the SBR system is the complex modulated QMF bank. The complex-valued representation permits modification of the subband samples without introducing excessive aliasing. SBR has proven its value in three waveform coders, and is already established on different markets. The number of applications using SBR is expected to increase rapidly in the near future.

6. REFERENCES

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

[2] International Standard ISO/IEC 11172-3:1993, "Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s – Part 3: Audio," ISO/IEC, 1993.

[3] International Standard ISO/IEC 13818-7:1997, "Information technology – Generic coding of moving pictures and associated audio information – Part 7: Advanced Audio Coding (AAC)," ISO/IEC, 1997.

[4] M. Dietz and S. Meltzer, "CT-aacPlus – a state-of-the-art Audio coding scheme," EBU Technical Review, July 2002, http://www.ebu.ch/trev_291-dietz.pdf.

[5] XM Satellite Radio, http://www.xmradio.com.

[6] S. Meltzer, R. Böhm and F. Henn, "SBR enhanced audio codecs for digital broadcasting such as "Digital Radio Mondiale" (DRM)," in 112th AES Convention, Munich, May 2002.

[7] ETSI TS 101 980 v1.1.1 (2001-09), "Digital Radio Mondiale (DRM); System Specification," ETSI, 2001.

[8] T. Ziegler, A. Ehret, P. Ekstrand and M. Lutzky, "Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm," in 112th AES Convention, Munich, May 2002.

[9] ETS 300 401, "Radio broadcasting systems; Digital Audio Broadcasting (DAB) to mobile, portable and fixed receivers," ETSI/EBU, 1997.

[10] International Standard ISO/IEC 14496-3:2001/FPDAM 1, "Bandwidth Extension," ISO/IEC, 2002.

[11] P. P. Vaidyanathan, "Multirate Systems and Filter Banks," Prentice Hall: Englewood Cliffs, NJ, 1993.

[12] H. S. Malvar, "Signal Processing With Lapped Transform," Artech House: Norwood, MA, 1992.

[13] K. Brandenburg, "Introduction to Perceptual Coding," in AES, Collected Papers on Digital Audio Bitrate Reduction, 1996.

[14] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C, The Art of Scientific Computing, Second Edition," Cambridge University Press, NY, 1992.