Audio Engineering Society
Convention Paper 5553
Presented at the 112th Convention
2002 May 10–13 Munich, Germany

Spectral Band Replication, a novel approach in audio coding

Martin Dietz[1], Lars Liljeryd[2], Kristofer Kjörling[2] and Oliver Kunz [1]

[1] *Coding Technologies, Nuremberg,Germany*

[2] *Coding Technologies, Stockholm, Sweden*

Correspondence should be addressed to Martin Dietz (diz@codingtechnologies.com)

ABSTRACT

Spectral Band Replication (SBR) is a novel technology which significantly improves the compression efficiency of perceptual audio codecs. SBR reconstructs the high frequency components of an audio signal on the receiver side. Thus, it takes the burden of encoding and transmitting high frequency components off the encoder, allowing for a much higher audio quality at low datarates. In December 2001, SBR has been chosen as Reference Model for the MPEG standardization process for bandwidth extension.

The paper will highlight the underlying technical ideas and the achievable efficiency improvements. A second focus will be a description of current and future applications of SBR.

INTRODUCTION

Back in 1997, Coding Technologies started research on a novel approach to improve the efficiency of digital audio coding algorithms. A team of Swedish researchers led by Lars Liljeryd had the idea to recreate (replicate) the high-frequencies of a decoded audio signal in an accurate way, mainly by the re-use of signal information obtained from the decoded base-band signal. Thus a new concept of redundancy coding in the frequency domain was developed. This concept led to what is today known as Spectral Band Replication technology (SBR). SBR is able to significantly improve the compression efficiency of audio codecs over a wide range of bitrates.

As SBR is an enhancement technology, it always needs an audio codec to hook upon. Two obvious candidates were identified: mp3 [1,2,6], as the worlds most popular coding format, and MPEG AAC [3,7], so far the most powerful waveform audio coding algorithm. After a year of successful cooperation with the Fraunhofer Institute for Integrated Circuits (FhG-IIS), several mp3 and AAC experts from FhG joined Coding Technologies in autumn 2000 to work on the SBR enhanced audio coding systems. Since then, several applications have adopted SBR. mp3PRO is the result of combining mp3 with SBR in a back- and forwards compatible way. mp3PRO is available since June 2001 and has in the meantime gained significant market acceptance. In January 2001, AAC+SBR has been adopted as audio codec for the Digital Radio Mondiale System. In December 2001, SBR has been successfully submitted to MPEG and became reference model of the MPEG-4 version 3 audio standardization process. The following chapters will describe in more detail the background of SBR, the technology behind SBR and applications for this novel technology.

BACKGROUND OF THE SBR DEVELOPMENT

The use of means to recreate missing or reduced high-frequencies in audio is nothing new, and has been used more or less frequent over the years in the recording industry. One of the most famous products was the "Aphex Aural Exciter" invented by Kurt Knoppel in the early 70's. Although Knoppel was not the first (or last) in this field, similar technologies can be found in some audio editing software tools today as well as in some audio processing hardware units. The main aim with such technologies is to add some "sparkle", "warmth" or high-end "enrichment" of the recorded vocals, speech, instruments, CD-master or broadcast feed, using typically non-linearities or "controlled" forms of high-frequency distortion. However, such technologies are performance limited by the way the high-frequencies are recreated. This limits their scope of use to just "touching-up" existing programme material, and making them unable to replicate major frequency ranges of a high quality programme material.

A related technology, "Harmonizing", was first successfully pioneered by the Eventide Clockworks "Harmonizer" in the early 70's, where programme material can be transposed in a "musically correct" manner. This is quite opposite to the "Bode Frequency Shifter" developed for electronic music applications by Harald Bode where harmonic relations are "non-musically" shifted (scrambled). Finally, the Ring-Modulator, one of the oldest "weird" sound effects units, is sometimes used in electronic music. Spin-off's from Harmonizer-like algorithms are the popular pitch-shift and time-stretch algorithms used in many audio workstation sound-editing software tools today.

The above-mentioned technologies or methods are sufficient for their intended use, but unsuitable for high quality SBR-like High-Frequency-Reconstruction (HFR) methods as introduced by Coding Technologies for audio coding. Therefore the development of new (patented) and more suitable algorithms was a prerequisite for the successful deployment of SBR. Strict demand was set on basic algorithm performance due to the unusual need to transpose signal components from the available (coded) baseband, by at least a factor three, without creating annoying artefacts.

Particular emphasis was put on recreating psychoacoustically accurate representations of the original high-band signal components without violating the well-known consonance-dissonance criteria. However compared to the original signal, the recreated high-band is not necessarily signal-coherent in a technical sense, but coherent in a mere psychoacoustic sense, a most important characteristic.

WHY SBR IMPROVES AUDIO CODING

Research on perceptual audio coders started about 20 years ago. Research on the human auditory system revealed that hearing is mainly based on a short-term spectral analysis of the audio signal. The so-called masking effect was observed: the human auditory system is not able to perceive distortions that are masked by a stronger signal in the spectral neighborhood. Thus, when looking at the short-term spectrum, the so-called masking threshold can be calculated. Distortions below this threshold are – in the ideal case - inaudible. Figure 1 illustrates in a very simplified way how the masking threshold can look like. Figure 1 also reveals that the distortion generated by the PCM format used on CD's is for all frequencies way below this threshold. However, since the PCM signal, representing the time domain, does not allow frequency dependent shaping of the quantization noise, a high SNR needs to be applied in order to achieve high quality sound across a large dynamic range. This also reveals the basic idea of perceptual audio coding: if it would be possible to shape the quantization noise in the frequency domain, compression of audio signals beyond the CD format must be achievable.

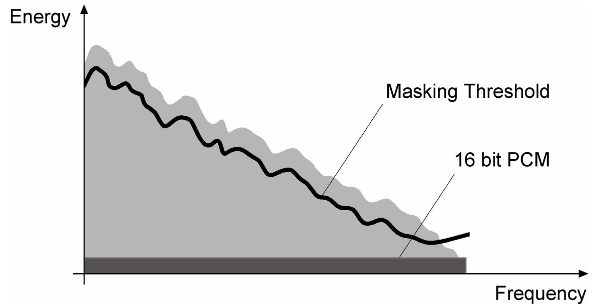

Figure 1: Spectrum and Masking Threshold

Research started on how to calculate the masking threshold ("psychoacoustic model") and how to process the audio signal in a way that only audible information resides in the signal. The ideal audio codec introduces quantization distortions that are exactly below the masking threshold (figure 2).

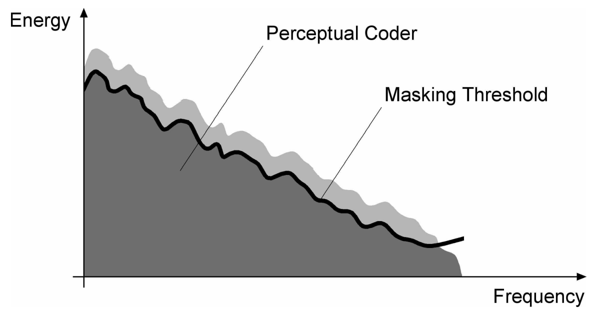

Figure 2: Ideal Perceptual Coding

This research led to todays waveform based perceptual audio codecs, such as MPEG Layer 2, AC-3, mp3, Atrac, PAC and MPEG AAC. All these codecs are based on the same, well-known principle, as shown in figure 3. The audio signal is transformed into the frequency domain by means of a filterbank or transform on a block-by-block basis. The resulting short-time spectrum is quantized in a way that the masking threshold calculated by the psychoacoustic model is not violated. The quantized spectrum gets coded and packed into a bitstream. The decoder performs the reverse steps.

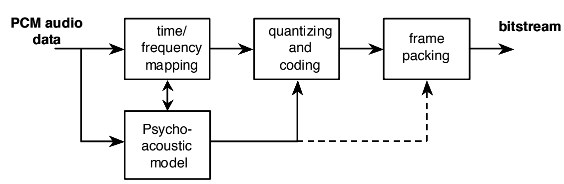

Figure 3: Perceptual waveform encoder

Although all the codecs mentioned follow the same principle, they still differ significantly in compression efficiency. The performance of such codecs depends mainly on two aspects: how precisely the masking threshold can be approached and how efficiently the quantized spectrum can be coded. AAC, for example, is, according to several tests, twice as efficient as Layer 2.

Nevertheless, traditional waveform coding has its limits. Having reached a performance like AAC it gets very hard to further increase compression. Figure 4 illustrates what happens if the bitrate is lowered significantly from the codecs reasonable operating point. The quantization noise significantly exceeds the masking threshold, thus generating audible artefacts.

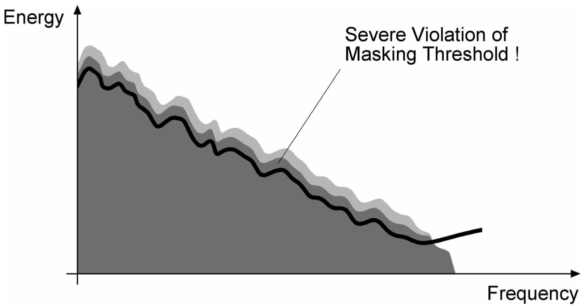

Figure 4: Waveform coding beyond its limits

Two main methods have been used so far to overcome this problem in perceptual waveform codecs. The most important one is to limit the audio bandwidth of the signal in or prior to the coding process. As there is no high frequency energy to be coded, more information is available for the remainder of the spectrum, resulting in a clean, but hollow sounding signal (figure 5). The other method, called intensity stereo, can only be used for stereo signals. In intensity stereo, only one channel and a panning information is transmitted instead of a left and a right channel. However, this is only of rather limited use for increasing compression efficiency, as in many cases the stereo image of the audio signal gets destroyed. A third method used is insertion of artificial noise. This method, however, works only with a rather limited class of audio signals.

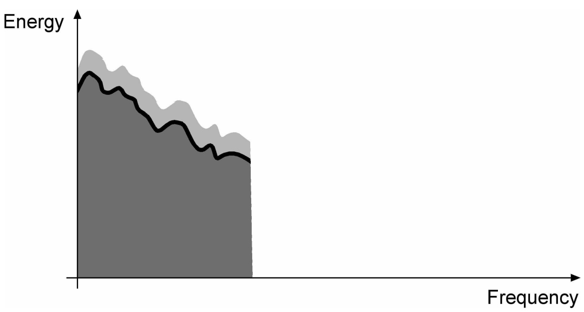

Figure 5: Limiting the audio bandwidth

At this point, the SBR technology comes on the scene. SBR deviates from the waveform coding principle towards a hybrid waveform/parametric method. It is based on the fact that in most cases there are large dependencies between the lower and higher frequency parts of an audio signal. Therefore, the high frequency part of an audio signal can be efficiently reconstructed from the low frequency part. Transmission of the high frequency part is therefore not necessary - only a small amount of SBR control data needs to be carried in the bitstream to guarantee an optimal reconstruction of the high frequencies. Figure 6 illustrates the first step in performing the SBR enhancement. The low frequency part is still coded by an ordinary waveform codec such as mp3 or AAC. The high frequency part, however, is generated by a high quality transposition algorithm.

As can be seen in figure 6, the mere generation of high frequency content is not at all sufficient for accurate high frequency reconstruction, since the reconstructed part does not reflect the spectral envelope of the original. Therefore careful adjustment of the spectrum is essential for the performance of the system. The adjustment is controlled by the SBR information carried in the bitstream and results in a correctly shaped high frequency part (figure 7).

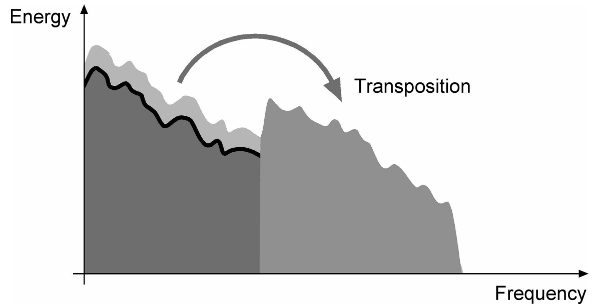

Figure 6: High frequency generation based on the waveform coded low frequency part

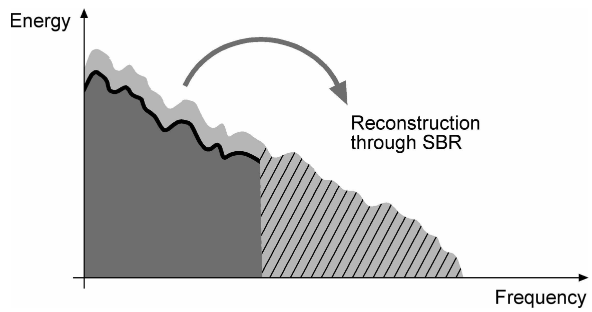

Figure 7: Spectrum after high frequency adjustment

To summarize, SBR enhanced codecs perform significantly better mainly because
a) SBR allows the recreation of the high frequencies using only a very small amount of transmitted side information. The high frequencies, which normally consume a significant amount of bits, do not need to be waveform coded anymore, resulting in a significant coding gain
b) the underlying waveform coder can run with a comparatively high SNR, as it is only responsible for the lower frequencies. It can even operate at the optimum sampling rate, which is usually different from the desired output sampling rate. SBR converts the waveform codec sampling rate into the desired output sampling rate.
Obviously there are signals where the reconstruction method does not deliver the desired results, e.g. when there is little relationship between the low and high frequency part. Care has been taken to equip SBR with additional tools so that such situations can be handled very well without losing compression efficiency.

HOW SBR WORKS

The SBR system is preferably used as a dual-rate system, with the underlying codec operating at half the original sampling-rate, while SBR operates at the original sampling rate. The following description will briefly explain the different parts in the encoder and decoder of the SBR system.

The SBR encoder works in parallel with the underlying core codec, albeit at a higher sampling-rate. Although SBR is mainly a post process in the decoder, important parameters are extracted in the encoder in order to ensure the most accurate high frequency reconstruction in the decoder. The encoder process is depicted in figure 8. The encoder estimates the spectral envelope of the SBR range for a time and frequency range/resolution suitable for the current input signal segments characteristics. The spectral envelope is estimated by a complex QMF analysis and subsequent energy calculation. The time and frequency resolutions of the spectral envelopes can be chosen with a high level of freedom, in order to ensure the best suited time frequency resolution for the given input segment. The envelope estimation needs to consider that a transient in the original, mainly situated in the high frequency region (for instance a high-hat), will be present to a minor extent in the SBR generated highband prior to envelope adjustment, since the highband in the decoder is based on the low band where the transient is much less pronounced compared to the highband. This aspect imposes different requirements for the time frequency resolution of the spectral envelope data, compared to ordinary spectral envelope estimation as used in other audio coding algorithms.

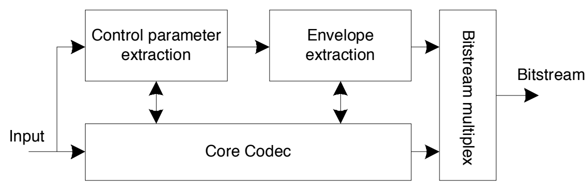

Figure 8: Basic block-diagram of the SBR encoder.

Apart from the spectral envelope, several additional parameters are extracted representing spectral characteristics of the input signal for different time and frequency regions. Since the encoder naturally has access to the original signal as well as information on how the SBR-unit in the decoder will create the high-band, given the specific set of control parameters, it is possible for the system to handle situations where the lowband constitutes a strong harmonic series and the high band, to be recreated, mainly constitutes random signal components, as well as situations where strong tonal components are present in the original highband without counterparts in the lowband, upon which the highband region is based. Furthermore, the SBR encoder works in close relation to the underlying core codec to assess which frequency range should be covered by SBR at a given time. The SBR data is efficiently coded prior to transmission by exploiting entropy coding as well as channel dependencies of the control data, in the case of stereo signals.

The control parameter extraction algorithms need to be carefully tuned to the underlying codec at a given bitrate and a given sampling rate. This is due to the fact that a lower bitrate, usually implies a larger SBR range compared to a high bitrate, and different sampling rates correspond to different time resolutions of the SBR frames. Different core codecs may also display different characteristics during bit constraints, i.e. more or less severe spectral band shut-downs etc.

The decoder is constituted of several different parts, as outlined in figure 9. It comprises a bitstream decoding module, a high frequency reconstruction (HFR) module, an additional high frequency components module, and an envelope adjuster module. The system is based around a complex valued QMF filterbank [4].

In the bitstream extraction module the control data is read from the bitstream and decoded. The time frequency grid is obtained for the current frame, prior to reading the envelope data from the bitstream, and decoding of the same. If a stereo signal is processed, suitable stereo decoding is performed according to the control data. Appropriate CRC-checks and error resilience algorithms are also included.

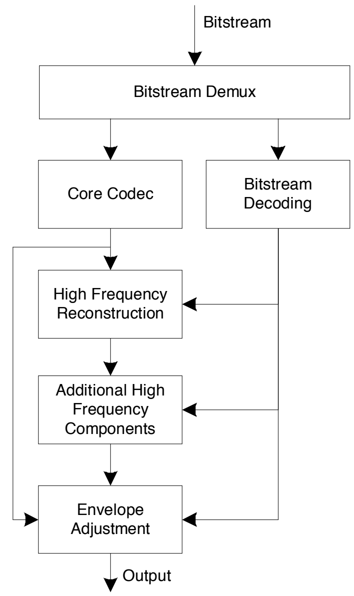

Figure 9: Basic block-diagram of the SBR decoder.

The underlying core decoder decodes the audio signal of the current frame (albeit at the lower sampling rate). The resulting frame of audio data is used for high frequency reconstruction by the HFR module. A possible output from the core decoder is displayed in figure 10. Here it is evident that the output is of low-pass character with a bandwidth of approx. 6kHz.

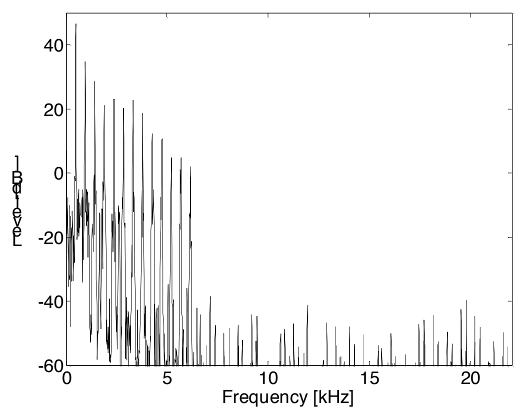

Figure 10: Spectrum of the output signal from the core decoder, for an input signal coded with AAC+SBR at 24kbps.

The decoded low-band signal is analyzed using a QMF filterbank. The high frequency reconstruction and envelope adjustment is subsequently performed on the subband samples of the QMF filterbank. The high frequencies are reconstructed from the low-band in a flexible way, based on the given control parameters. Furthermore, the reconstructed highband is adaptively filtered on a subband channel basis according to the control data to ensure the appropriate spectral characteristics of the given time/frequency region.

The HFR can be done in several ways dependent on the characteristics of the signal being processed. The output from such a transposition based high frequency reconstruction is depicted in figure 11. Here it is evident that the frequency range of the signal has been extended to approx. 15kHz. There are no algorithmic constraints preventing the HFR module to extend the frequency range even higher. However, for this example 15kHz was deemed sufficient.

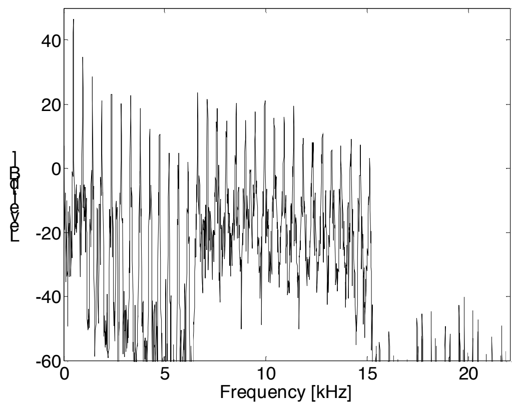

Figure 11: Spectrum of the output signal from a transposition based SBR module, given the input signal displayed in figure 11

The reconstructed highband is subsequently envelope adjusted according to the time/frequency resolution obtained from the encoder. The output signal is enhanced with additional signal components in order to ensure the correct spectral characteristics of the final output.

The low band decoded by the underlying decoder is added to the reconstructed and envelope adjusted highband, giving the output signal. Since the lowband has a sampling rate of half of the highband it is up-sampled during the course of the SBR process. The final output, after envelope adjustment, is depicted in figure 12. It is evident that the SBR module successfully has recreated a high band very similar to that of the original, fig 13.

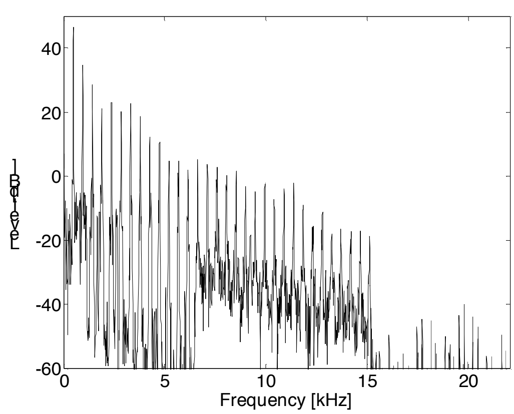

Figure 12: Spectrum of the output signal after envelope adjustment, coded with AAC+SBR at 24kbps.

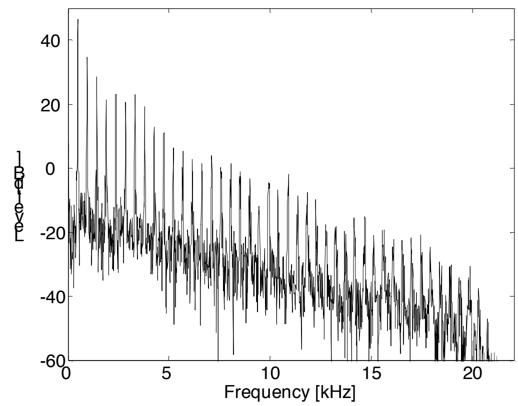

Figure 13: Spectrum of the original signal.

SBR APPLICATIONS

SBR enhanced audio codecs are already in use in several applications, some of which are listed hereafter.

mp3PRO

In January 2001, Coding Technologies and its licensing partner Thomson Multimedia presented mp3PRO, the combination of mp3 and SBR, for the first time. In June 2001, a free demo software has been made available. Since then several software applications integrating mp3PRO have been released (e.g. from Ahead, Steinberg, Magix, ….). First hardware products are expected in Q2/2002.
mp3PRO offers significantly improved performance compared to mp3 and outperforms competitive codecs according to independent tests.
mp3PRO is back- and forward compatible with mp3. Not only will any mp3PRO decoder decode mp3 content, all mp3 players will be able to decode mp3PRO bitstreams, although without the quality improvement achieved through SBR.

Figure 14 shows the integration of mp3 and SBR . Basically, SBR can be seen as pre/postprocessing around the existing mp3 modules. Even the bitstream format remains unchanged, as the SBR data can be carried in the ancillary data field.
More detailed information about mp3PRO can be found in [4].

Digital Radio Mondiale (DRM)

The DRM consortium has been founded in 1998 with the goal to standardize a new digital system for medium, long- and shortwave transmissions. The system should deliver significantly better audio and reception quality than ordinary analog AM. At the same time the existing AM channel spacing should be used. In addition, reuse of existing transmitter equipment should be possible to guarantee a cost efficient introduction of the new system.

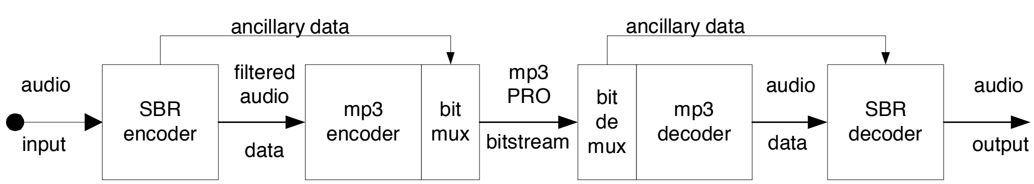

Figure 14: mp3PRO block diagram

Depending on the wavelength and the region, the channel spacing in the AM bands is either 9 kHz or 10 kHz. The propagation conditions, which depend very much on the wavelength, determine the spectral efficiency achievable. Extensive tests have resulted in an efficiency in the range of 1.5 to 3 bit/Hz. As a result, the bitrate available in one transmission channel is in the range of 13…30 kbit/sec. Normal operation will use bitrates between 20 and 25 kbit/sec.
Delivering high quality audio at such a low bitrate asks for the most powerful source coding scheme available. Consequently, AAC+SBR, the combination of MPEG-AAC and SBR has been chosen as audio coding method. Tests done within the DRM consortium clearly show the superiority of this coding system compared to previous state of the art. More details about DRM and the listening test can be found in [5,8,9] and at the DRM web site www.drm.org.
DRM finalised the specification of the system in January 2001. It is recommended by the ITU [11] and standardised by ETSI [10]. First regular transmissions and receiver products are expected in 2003.

MPEG

Because of its outstanding compression performance, the combination of AAC and SBR is of high interest not only for broadcasting, but also for audio and audio/video applications like streaming over the internet, streaming and delivery over mobile networks and storage in portable devices. For a lot of these applications, open standards like MPEG play an important role. They usually provide state-of-the-art algorithms and guarantee interoperability and accessibility of the technology.
As a consequence, MPEG issued a call for proposals in January 2001, asking for technologies that could further enhance the compression efficiency of the MPEG-4 audio coding algorithms (in particular AAC). SBR was submitted as a proposal; a competitive test was performed by MPEG. AAC+SBR did not only show an excellent performance, but already met the acceptance criteria set up for final acceptance of the technology in the standard. As a result, SBR became Reference Model 0 for version 3 of the MPEG-4 standard.
Figure 15 shows the result of the test at 24 kbps mono at the T-Nova test site. The test was done with experienced listeners and the well proven MUSHRA test method. As test items, the "usual" MPEG test items were used (e.g. harpsichord, glockenspiel, pitchpipe, male german speech, castagnettes, a.s.o). Till today those items are known to be very critical for audio coding. AAC+SBR shows a vast improvement compared to AAC, even when compared to the higher bitrate AAC (30 kbps). Moreover AAC+SBR shows a very good absolute performance at this very low bitrate.

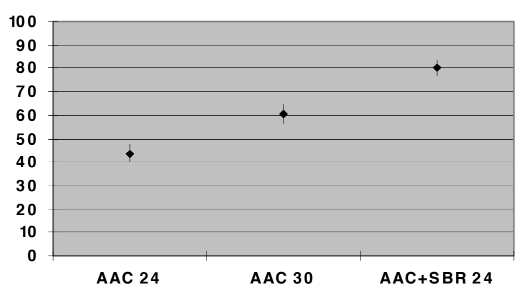

Figure 15: MPEG test result at 24 kbps mono (test site: T-Nova)

More detailed results are given in figure 17. Table 1 explains the acronyms for the excerpts.

Es01	Suzanne Vega
es02	Male German Speech
es03	Female English Speech
sc01	Haydn Trumpet Concert
Sc02	Meistersinger
Sc03	Spot
Si01	Harpsichord
Si02	Castagnettes
Si03	Pitchpipe
Sm01	Bagpipe
Sm02	Glockenspiel
Sm03	Plucked Strings

Table 1: The MPEG excerpts

The stereo test results are shown in figure 16 and Figure 18. Again, AAC+SBR shows a significant gain over AAC and a very good absolute grading of more than 85 at a bitrate of 48 kbit/sec. For both tests it should be noted that the comparison was done against MPEG-4 AAC. A comparison to the yet much more popular MPEG-2 AAC, which is the AAC version available on the market today, would show an even higher improvement through SBR (which operates based on MPEG-2 AAC). Since AAC was so far deemed to be state-of-the-art audio compression, it can be concluded from the this test that AAC+SBR is the new performance leader in audio coding.

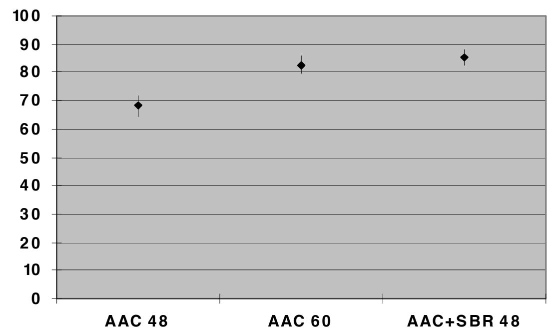

Figure 16: MPEG test result at 48 kbps stereo (test site: T-Nova)

The MPEG-4 version 3 standard will reach Committee Draft (CD) status in May 2002 and the Final Draft International Standard (FDIS) status in March 2003. The improved audio coding standard is expected to be adopted for several of the above mentioned applications.

CONCLUSIONS

Spectral Band Replication is a novel technology that combines traditional audio coding with the capabilities of high quality high frequency reconstruction methods. Through the use of SBR in mp3PRO, the compression efficiency of mp3 could be significantly improved while remaining compatible with the widespread mp3 format. mp3PRO will succeed mp3 in all application areas.
The combination of AAC and SBR offers the most powerful audio compression available today. As such it is most suited for digital broadcasting and streaming/delivery over networks with limited resources. AAC+SBR is already used in the market place and is the Reference Model in the MPEG-4 standardization.
Both mp3PRO and AAC+SBR are available today in soft- and hardware, several applications have already been launched. It is expected that the number of applications using SBR will increase significantly in the near future.

REFERENCES

[1] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 11172-3 "Information technology – Coding of moving pictures and associated audio for digital storage media up to about 1,5 Mbit/s", Part 3: Audio, 1991.

[2] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 13818-3 ”Information technology – Generic Coding of Moving Pictures and Associated Audio:”, Part 3: Audio, 1994.

[3] ISO/IEC JTC1/SC29/WG11 MPEG, International Standard ISO/IEC IS 13818-7 ”Information technology – Generic Coding of Moving Pictures and Associated Audio:”, Part 7: Advanced Audio Coding (AAC), 1997.

[4] Per Ekstrand, Andreas Ehret, Manfred Lutzky and Thomas Ziegler. Enhancing mp3 with SBR: Features and Capabilities of the new mp3PRO Algorithm. In 112th AES Convention, Munich, May 10-May 13, 2002.

[5] Stefan Meltzner, Fredrik Henn and Reinhold Böhm. SBRenhanced audio codecs for digital broadcasting such as “Digital Radio Mondial” (DRM). In 112th AES Convention, Munich, May 10-May 13, 2002.

[6] K. Brandenburg, G. Stoll, Y.F. Dehery, J.D. Johnston, L.v.d. Kerkhof, E.F. Schröder: "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio", JAES 42, 780-792

[7] Bosi, Brandenburg, Quackenbush, Fielder, Akagiri, Fuchs, Dietz, Herre, Davidson, Oikawa, “ISO/IEC MPEG-2 Advanced Audio Coding”, 101st AES Convention, Los Angeles, November 1996

[8] DRM Source Coding Group, 2001: Report on Subjective Listening Tests of SBR-LC, an AAC-based Audio Bandwidth Widening Tool

[9] Jonathan Stott, 2001. DRM – Key Technical Features, Reprint from EBU Technical Review No. 286 (March 2001)

[10] ETSI TS 101980 v1.1.1 (2001-09), Digital Radio Mondiale (DRM); System Specification

[11] Recommendation ITU-R BS.1348-1 (02/01): Service requirements for digital sound broadcasting at frequencies below 30 MHz

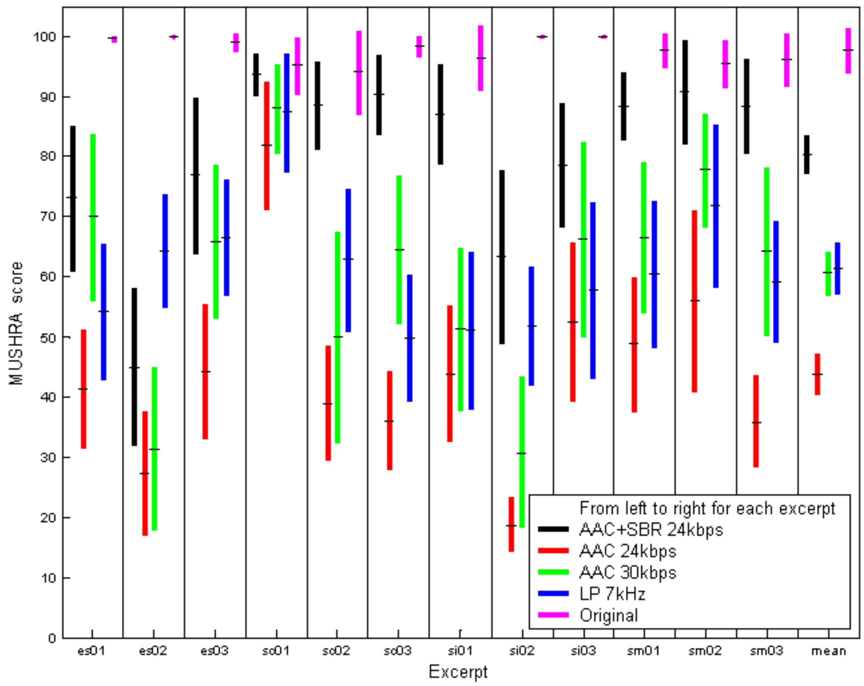

Figure 17: MPEG test results per item, 24 kbps mono (test site: T-Nova)

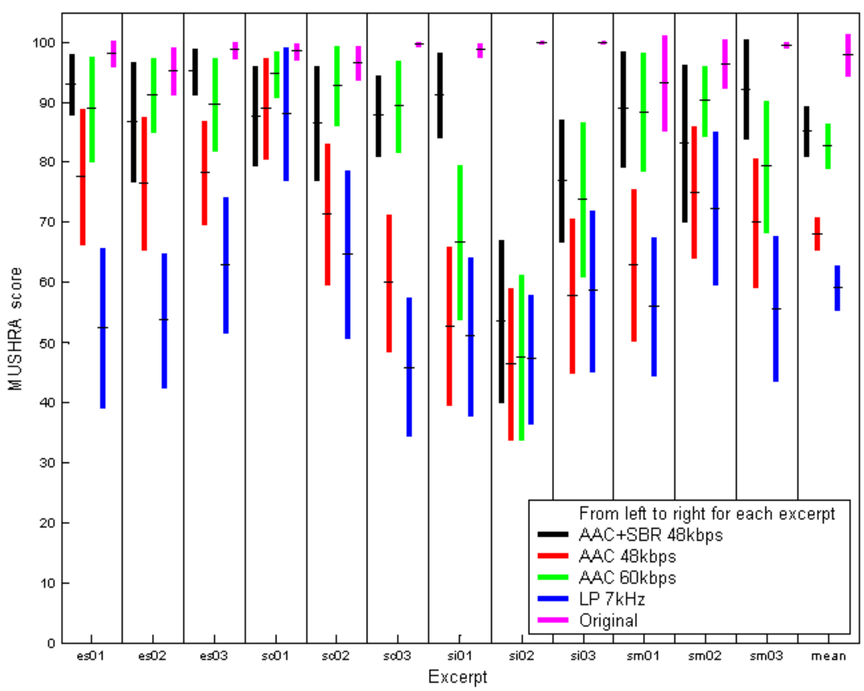

Figure 18: MPEG test results per item, 48 kbps (test site: T-Nova)