Enhancing Perceptual Audio Coding through Spectral Band Replication

Martin Dietz and Martin Wolters
Coding Technologies, Nuremberg, Germany; Email: {diz|wol}@codingtechnologies.com

Introduction

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 data rates.

This 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.

Why SBR improves Audio Coding

Today's popular audio codecs such as mp3[1] and MPEG AAC[2] are so called waveform audio coding algorithms. The basic principal is the same for all such codecs: The encoder estimates the masking threshold and tries to shape the quantization noise in the frequency domain to be below that threshold. Whereas estimating the masking threshold is independent of the specific coding standard, the ability to shape the quantization noise is largely dependent on the bitstream format as defined in a specific coding standard. For example, several tests have shown that AAC is twice as efficient as Layer 2.

Nevertheless, traditional waveform coding has its limits. Having reached a performance like AAC it becomes difficult to further increase compression. This is specifically true for bitrates at or below 64 kbps for a stereo file at 44.1 kHz sampling rate. Until today, the most important method in such situation was limiting the audio bandwidth of the signal prior to the coding process. As there is no high frequency energy to the coding process, more information is available for the remainder of the spectrum, resulting in a clean, but hollow sounding signal.

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 (~ 3kbps) needs to be carried in the bitstream to guarantee an optimal reconstruction of the high frequencies. Figure 1 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.

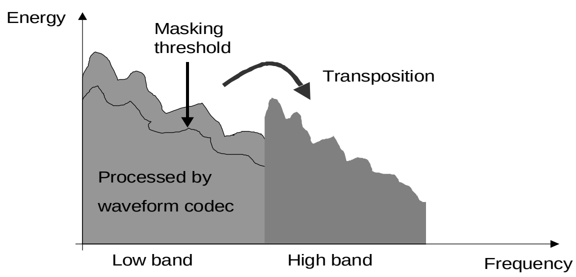

Figure 1: High frequency generation based on the waveform coded low frequency part

As can be seen in figure 1, the mere generation of high frequency content is insufficient 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 2).

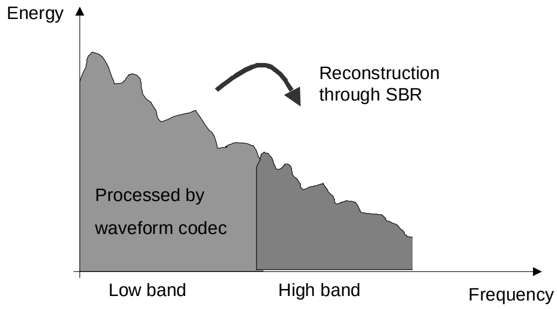

Figure 2: Spectrum after high frequency adjustment

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 of the SBR system. See this AES preprint[3] for a more detailed description.

The SBR encoder works in parallel with the underlying core codec. 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. These parameters include the frequency range of the core codec, a varying time/frequency grid to be used by the envelope adjuster, as well as other side chain data to be used by the SBR decoder. This data is then entropy coded and combined with the bitstream from the core codec.

The decoder process is depicted in figure 3. It comprises a bitstream decoding module, a high frequency generation module, an additional high frequency components module, and an envelope adjuster module. The system is based around a complex valued QMF filterbank.[4] The SBR decode process follows the core decoder and all operations are guided by parameters specified and transmitted by the encoder.

The HF generator transposes data from the low into the high band and applies an inverse filter to the generated high frequency content. The additional HF components module is used for audio signals that do not have a strong correlation between the low- and high band. Thus, it is possible for the system to handle situations

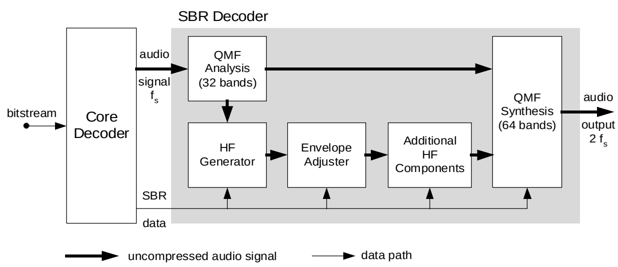

Figure 3: Basic block diagram of the SBR decoder

1 K. Brandenburg, et al.: "The ISO/MPEG-Audio Codec: A Generic Standard for Coding of High Quality Digital Audio", JAES 42, 780-792

2 Bosi, et al.: “ISO/IEC MPEG-2 Advanced Audio Coding”, 101st AES Convention, Los Angeles, November 1996

3 Martin Dietz, et al.: “Spectral Band Replication, a novel approach in audio coding”, 112th AES Convention, Munich, May 2002.

4 Per Ekstrand: “Bandwidth Extension of Audio Signals by Spectral Band Replication”, IEEE Benelux Workshop on Model based Processing and Coding of Audio (MPCA-2002), Leuven Belgium, Nov 15, 2002

where the low band constitutes a strong harmonic series but the high band mainly constitutes random signal components, and vice versa. Finally, the envelope adjuster compares the envelope of the generated data to envelope parameters specified by the encoder and adjusts frequency dependent gain parameters accordingly.

How SBR sounds

The combinations of SBR with popular waveform coding algorithms such as mp3 (called "mp3PRO") and AAC (called "aacPlus") were compared against other audio coding formats in several tests. A well respected, independent test was performed by the European Broadcasting Union (EBU) and was published in September 2002[5]. The test was conducted using the EBU MUSHRA test method[6]. Figure 4 shows the results at 48kbps stereo. It clearly shows the advantage of adding SBR to an existing waveform codec. The scores for mp3 can be increased from about 40 to well above 60, the scores for AAC from above 60 to about 80.

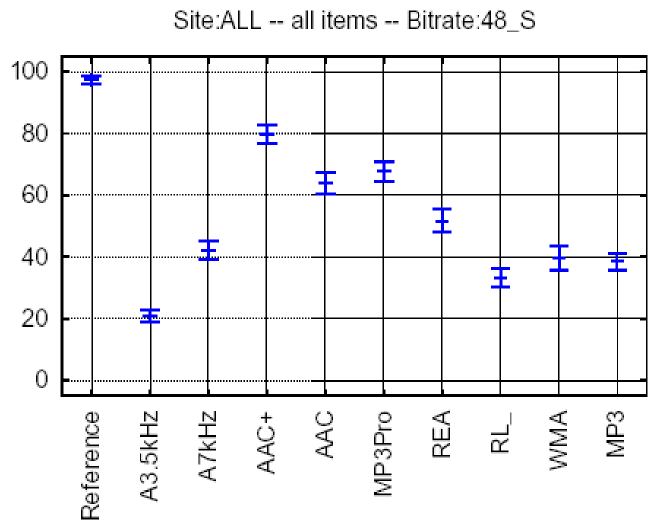

Figure 4: Results of the EBU listening test for 48kbps stereo

Note, that a test score above 80 is considered "excellent" quality, whereas items between 20-40 are considered "poor" quality. In March of 2003 several MPEG contributors also performed a similar listening test using the EBU MUSHRA test method. Figure 5 illustrates a subset of these test results and it again shows the improvements achieved by adding SBR to an existing waveform codec.

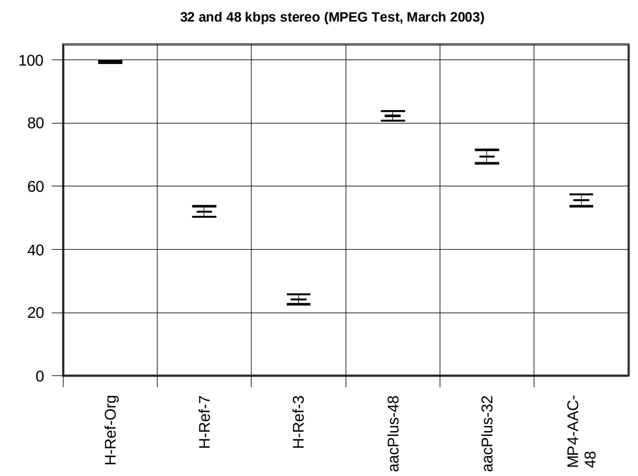

Figure 5: Test results from MPEG contributors

5 BPN 049: "The EBU Subjective Listening Tests on Low Bitrate Audio Codecs", EBU Project Group B/AIM, September 2002

6 ITU-R Recommendation BS.1534 (June 2001): Method for the subjective assessment of intermediate quality level of coding systems

SBR Applications

mp3PRO

In January 2001, Coding Technologies and its licensing partner Thomson Multimedia introduced mp3PRO[7], the combination of mp3 and SBR. Since then several software applications and hardware products integrating mp3PRO have been released. (See http://www.mp3PROzone.com for a detailed list)

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, however without the quality improvement achieved through SBR.

Digital Radio Broadcasting

Digital radio broadcasting systems such as XM Radio and Digital Radio Mondiale (DRM) are among the newest consumer audio applications. They offer a more efficient use of the limited frequencies available for broadcasting while enabling detailed control of the audio quality.

Both XM Radio as well as DRM chose aacPlus as the basis for their systems since it guarantees highest audio quality at lowest bitrates. XM Radio already counts more than 350,000 subscribers. The DRM standard has already been finalized, is recommended by the ITU, and was standardized by ETSI.

Open Standards: MPEG 4 High Efficiency AAC Profile

Open standards like MPEG play an important role for multimedia applications like broadcasting, streaming over the internet, delivery over mobile networks and storage in portable devices. They usually provide state-of-the-art algorithms and guarantee both interoperability and accessibility of the technology.

In March 2003 MPEG finalized the MPEG 4 High Efficiency AAC Profile standard which is based on aacPlus. After detailed listening tests and technical evaluations the group decided that aacPlus offers the best audio performance at lower bitrates.

Conclusions and the Future of SBR

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.

The combination of AAC and SBR, aacPlus, offers the most powerful audio compression available today. As such, it is best suited for digital broadcasting and streaming/delivery over networks with limited resources. aacPlus is already used in the market place and is part of the MPEG-4 standard.

It is expected that the number of applications using SBR will increase significantly in the near future. Ongoing evaluations of this technology by independent organizations will likely result in further standardization of SBR, new combinations of core codecs with SBR (incl. speech codecs), and new application areas in broadcasting and for mobile devices.

The basic concept of combining waveform with parametric coding will be extended to other areas beyond high-frequency reconstruction such as parametric stereo coding. In such a scenario the SBR encoder limits the number of channels that need to be processed by the core codec (e.g. mono instead of stereo) and reconstructs the required output channels in the decoder.

The combination of existing core codecs with a system like SBR is a flexible and efficient method to provide further quantum leaps in audio coding. The SBR system already provides the required groundwork and infrastructure and will enable further extensions.

7 Per Ekstrand et al.: "Enhancing mp3 with SBR:" In 112th AES Convention, Munich, May 10-May 13, 2002.